

Investigación y desarrollo de nuevos métodos moleculares y bioinformáticos en salud pública para la detección de microorganismos y el estudio de la respuesta inmune humoral

Tesis Doctoral
Guillermo Hugo López Campos
(Licenciado en Ciencias Químicas. UAM)

Director
Dr. Fernando Martín Sánchez

Tutor
Dr. Javier Díaz Nido

ÍNDICE

Índice	i
Agradecimientos	iii
Abstract / Resumen en Inglés	2
1. Introducción	6
1.1. La biología molecular para la resolución de problemas en salud pública	8
1.2. Problemas de interés en salud pública susceptibles de ser analizados mediante técnicas moleculares	10
1.2.1. Detección de agentes de potencial uso en bioterrorismo	10
1.2.2. Identificación de agentes causantes de patologías respiratorias – Adenovirus	12
1.2.3. Infección por VIH y subtipos	14
1.3. Micromatrices de material biológico	15
1.3.1. Antecedentes	15
1.3.2. Metodología de trabajo con micromatrices	17
1.3.3. Bioinformática asociada con las micromatrices	18
1.3.4. Aplicaciones	20
2. Objetivos	22
3. Materiales y métodos	26
3.1. Obtención y tratamiento de muestras	28
3.1.1. Muestras de agentes infecciosos de interés en bioterrorismo	28
3.1.2. Muestras de virus respiratorios	30
3.1.3. Muestras para el análisis de la respuesta inmune humoral	30
3.2. Reacciones de amplificación y marcaje	31
3.3. Diseño de cebadores y sondas	33
3.4. Funcionalización de superficies y fabricación de las micromatrices	34
3.5. Reacciones de Hibridación DNA:DNA sobre micromatrices	34
3.6. Reacciones de extensión de cebador sobre micromatrices	36
3.7. Reacciones Antígeno-Anticuerpo sobre micromatrices	37
3.8. Escaneado, análisis de las imágenes y análisis estadístico de los datos	37
3.9. Secuenciación y análisis filogenético	37
4. Resultados	38
4.1. Sistemas de detección de patógenos humanos basados en micromatrices de material biológico	40
4.1.1. Fabricación de las micromatrices	40
4.1.1.1. Funcionalización y análisis de superficies	40
4.1.1.2. Diseño de sondas	43
4.1.2. Sistema para el análisis de reacciones de PCR “ <i>in Silico</i> ” y pretratamiento de datos para el diseño de sondas de ácidos nucleicos. BUSSUB	45
4.1.3. Marcaje fluorescente de las muestras	47
4.1.4. Detección de agentes infecciosos de interés en bioterrorismo	48
4.1.4.1. Diseño y generación de un DNA quimérico de variola minor	48
4.1.4.2. Reacciones de amplificación de los organismos de interés	51
4.1.4.2.1. Reacciones para la detección del género Burkholderia	51
4.1.4.2.2. Reacciones para la detección del género Brucella	53

4.1.4.3. Sondas para agentes susceptibles de ser empleados en bioterrorismo	55
4.1.4.4. Validación de los ensayos para la detección de patógenos de interés en bioterrorismo mediante ensayos de hibridación	59
4.1.4.4.1. Agua contaminadas	61
4.1.4.4.2. Muestras problema	62
4.1.4.5. Validación de de los ensayos para la detección de patógenos de interés en bioterrorismo mediante ensayos de extensión de primer .	64
4.1.5. Detección de adenovirus respiratorios	67
4.1.5.1. Sondas para la detección de adenovirus humanos implicados en patologías respiratorias	67
4.1.5.2. Validación de de los ensayos para la detección de virus respiratorios mediante el análisis de muestras clínicas	67
4.2. Análisis cuantitativo de la respuesta inmune humoral frente a diferentes subtipos de VIH empleando micromatrices de material biológico	70
4.2.1. Cuantificación de las imágenes, control de calidad y normalización de los datos (preprocesamiento y transformación)	71
4.2.1.1. Cuantificación	71
4.2.1.2. Preprocesamiento – Control de calidad	72
4.2.1.3. Preprocesamiento – Normalización	72
4.2.2. Determinación de positividad de las reacciones	76
4.2.3. Comparación entre sueros positivos, negativos y comparaciones entre subtipos en los ensayos con micromatrices	77
4.2.4. Comparación de entre resultados obtenidos mediante técnicas ELISA y micromatrices	84
4.2.5. Análisis de perfiles de reactividad con micromatrices	86
5. Discusión	90
5.1. Análisis de las superficies para la fabricación	92
5.2. Métodos de asistencia para el diseño de sondas y recuperación de secuencias	94
5.3. Sistemas de detección de microorganismos patógenos mediante micromatrices	95
5.3.1. Detección de patógenos de interés en bioterrorismo	95
5.3.2. Sistemas de detección de virus respiratorios	97
5.4. Análisis de la respuesta inmune humoral frente diferentes subtipos de VIH	98
5.4.1. Desarrollo de estrategias de control de calidad y análisis de datos	99
5.4.2. Comparaciones genética vs. reconocimiento humoral	100
5.4.3. Comparación resultados ELISA vs. Micromatrices	103
5.4.4. Caracterización de perfiles antigénicos mediante ensayos de “afinidad” ..	104
5.4.5. Interpretación de los resultados de afinidad frente a los resultados empleando los péptidos independientemente	105
6. Conclusiones	106
7. Resumen final	110
8. Referencias Bibliográficas	116
9. Anexo I - Publicaciones)	132

Agradecimientos

En primer lugar me gustaría agradecer a mi director de tesis Dr. Fernando Martín Sánchez por haberme dado la oportunidad de formar parte de su grupo, en el que he podido desarrollar este trabajo, por su paciencia y su liderazgo durante la realización de esta tesis. Gracias Fernando por haber sido no solo el director de esta tesis y haberme ayudado a crecer científicamente si no también por haberme apoyado, haberme hecho crecer personalmente y haber sido un amigo a lo largo de estos años.

Gracias también al Dr. Javier Díaz Nido por haber sido el tutor de esta tesis en la Universidad Autónoma.

Gracias al Instituto de Salud Carlos III en cuyas instalaciones se ha desarrollado este trabajo. Gracias también al Ministerio de Defensa, a través de la Fábrica Nacional de La Marañosa por financiar parte de la investigación recogida en esta tesis.

Muchas gracias al Servicio de Virología dirigido por la Dra. Pilar Pérez Breña, con quien se colaboró en la investigación con los adenovirus, en especial a las Dras Mayte Coiras y María Rosa López con quienes fue un auténtico placer trabajar y aprender durante esta colaboración. Gracias Mayte por tu paciencia.

Gracias también al Servicio de Diagnóstico y Referencia de Retrovirus y Papilomavirus dirigido por el Dr. Alfredo García Saiz con quien se colaboró en la investigación de la respuesta frente a la infección por VIH. Gracias al Dr. Luís García Albert por las enriquecedoras charlas que hemos tenido a lo largo de estos años.

Muchas gracias a los miembros del Laboratorio de Espiroquetas y Patógenos Especiales, en especial a Pedro, Raquel, Horacio y Manuela por haberme escuchado y ayudado cuando ha habido problemas y por su magnífica colaboración siempre, del Laboratorio de Arbovirus y Enfermedades Emergentes, Antonio, M^a Paz y Giovanni por su colaboración, y del Laboratorio de Biología Molecular de La Fabrica Nacional de La Marañosa, en especial gracias a Óscar y Daniel por una colaboración fluida y fructífera.

A todos los miembros, pasados, presentes y de adopción de BIOTIC, por haberme soportado todos estos años, por haber hecho que tantas horas por allí hayan merecido la pena laboral y lo que es más importante, personalmente. Muchas gracias a la Dra. Victoria López Alonso por haberme apoyado, por haberme empujado y por estar siempre ahí con una sonrisa, una buena palabra y con un empujón cuando me he quedado atascado, gracias de verdad Vicky!. Muchas gracias a Javi, por haberme proporcionado durante muchos años un punto de vista diferente de la vida y por haber padecido conmigo los rigores de la alimentación en el ISCIII. A “Juan Pe”, por su paciencia con mis ideas “esto es sencillito” y por su colaboración en BUSSUB. A Inma, “mi querida Padawan” por todo lo que compartimos juntos en el labo. A Óscar por la estadística y lo que compartimos en los paseos al Clínico. A Isa por tantos años juntos y por todas las ayudas y correcciones del inglés que ha hecho. A Candi, por todos estos años de amistad. A

Santi, Nati, Carmen, Benja, Ángeles por los años pasados juntos, y a “los nuevos” Kiko, M^a Carmen (por no parecer nueva y por su ayuda en el labo), Laura, Sandra y Jorge.

A “la banda del bus”, Carlota, M^a José, David, Ana, Esther... y de los que me olvide por hacer las esperas y viajes más agradables. Gracias en especial a Carlota por haber estado ahí cuando no la esperaba y por “pesada”.

A los Pepes, Gutiérrez y García, por haber confiado en mí, haberme dado una oportunidad y haberme enseñado que siempre seré “young & unexperienced”.

A mi familia y amigos, los que siempre habéis estado ahí confiando en mí, apoyándome y ayudándome. Papá, Mamá, Lulu, “Nani”, “Peich”, el clan del S. Juan, Fran, Jose, Pablo, Pedro, Santi, Sergio, los que no son del S. Juan, Richi, Diego... No hay palabras para daros las gracias por todo lo que habéis hecho por mí.

A Nuria, por todos estos años juntos y en especial por los últimos meses de paciencia conmigo y con mi “rincón”.

Por último, agradecer también a todos los que no he citado pero habéis estado ahí.



**Departamento de Biología
Molecular
Facultad de Ciencias
Universidad Autónoma de Madrid**



Investigación y desarrollo de nuevos métodos moleculares y bioinformáticos en salud pública para la detección de microorganismos y el estudio de la respuesta inmune humoral

Tesis Doctoral
Guillermo Hugo López Campos
(Licenciado en Ciencias Químicas. UAM)

Director
Dr. Fernando Martín Sánchez

Tutor
Dr. Javier Díaz Nido

Abstract

Abstract:

Microarray technologies have widely spread and have become a common tool in genomic applications and studies. Although they have been mostly used in gene expression analyses, their high throughput capabilities make them a valuable technology to be applied in other fields such as clinical diagnosis or microbiology. Recent events such as the Anthrax alerts after Sept. 11th, or the currently active influenza A/H1N1 epidemics have shown the necessity of fast and accurate techniques for the detection and identification of infectious diseases, and microarrays are a good candidate to be used in this context. The work presented hereby addresses the issue of how best microarray and bioinformatics technologies (Lopez-Campos *et al.*, 2008) can be jointly applied for the study and detection of emerging infectious diseases of interest in public health as well as for the study of the humoral immune response generated after the infection by different subtypes of VIH.

Development of microarray based techniques for the detection of pathogens includes the optimization of microarray manufacture and probe design processes. Poly-l-lysine and isothiocyanate modified-glass provide good surface properties for the attachment of oligonucleotide probes and other relevant molecules. A major topic for microarray success is probe design; therefore, it is necessary to develop tools that support this complex process. BUSSUB has been developed as a tool that helps in the selection of the sequence areas of interest for the design of probes as well as for performing electronic PCR.

Microarrays have also been designed and tested showing high sensitivity and specificity for the simultaneous detection of bioterrorism class A and B agents. To test the feasibility of the application of microarray technologies in public health laboratories, a microarray for the detection and characterization of human respiratory adenoviruses was developed. It was also successfully assayed with clinical samples from patients with respiratory infections. This study shows that microarrays can be effectively used in clinical environments and may become a powerful method for the study and diagnosis of infectious diseases (Lopez-Campos *et al.*, 2007).

Lastly, protein microarrays have been used to study and characterize the humoral response against different HIV subtypes. This approach shows that it is possible to analyse simultaneously in a quantitative way the reactivity of sera against HIV antigens coming from different HIV subtypes. This analysis enabled the characterization and identification of different response profiles against the different subtypes. These results facilitated also the comparison of microarrays with ELISA tests, showing differences between the results obtained with each of the technologies. The simultaneous detection of different dilutions was also useful for the analysis of the humoral response taking into account the affinity of the sera against the different immobilised antigens (Lopez-Campos *et al.*, 2006).

1. Introducción

1.1 La biología molecular para resolución de problemas en salud pública

Recientes acontecimientos tales como el brote de gripe A H1N1 (Coker, 2009), el brote de gripe aviar H5N1 del año 2005 en el sudeste asiático (Parry, 2005), el brote en el año 2003 del Síndrome Respiratorio Agudo Grave (SARS) causado por un coronavirus (Drosten *et al.*, 2003; Holmes, 2003; Wang *et al.*, 2003) o el envío en 2001 de cartas contaminadas con esporas de *Bacillus anthracis* (Dewan *et al.*, 2002), han mostrado la importancia de contar con sistemas diagnósticos avanzados que permitan la detección e identificación de nuevos agentes infecciosos y enfermedades emergentes en salud pública de una forma rápida y segura.

Los avances realizados en los últimos años en las técnicas y conocimientos en la biología molecular están sirviendo de soporte para el desarrollo de nuevas aproximaciones en el campo de la microbiología clínica y la salud pública. Estos avances están permitiendo nuevos enfoques y el desarrollo de nuevas técnicas diagnósticas moleculares. La extensión de las técnicas moleculares para la identificación de agentes patógenos presenta numerosas ventajas en comparación con las técnicas clásicas de identificación fenotípica. Las técnicas moleculares permiten la realización de identificaciones avanzadas gracias a las cuales se puede llegar a discriminar entre cepas dentro de una misma especie o subespecie. Algunos ejemplos de este tipo de transferencia de tecnología y conocimientos son la aplicación diagnóstica de técnicas clásicas de PCR (Polymerase Chain Reaction) (Mullis *et al.*, 1986) o sus variaciones como RAPD (Random Amplified Polymorphic DNA) (Haase *et al.*, 1995; Mazurier *et al.*, 1992; Shangkuan *et al.*, 2001), AFLP (Amplified Fragment Length Polymorphism) (Pyrce *et al.*, 2008), RFLP (Restriction Fragment Length Polymorphisms) (Thomas *et al.*, 2003), PFGE (Pulsed Field Gel Electrophoresis) (García Del *et al.*, 2002) o técnicas basadas en la realización de inmunoensayos.

La genómica se ha desarrollado desde los años 90 gracias a los desarrollos en secuenciación y se puede considerar la obtención del genoma completo del organismo *Haemophilus influenzae* (Fleischmann *et al.*, 1995) como el comienzo del periodo conocido como era genómica. La conclusión del Proyecto Genoma Humano (Lander *et al.*, 2001; Venter *et al.*, 2001) representa su hito más importante. Con el desarrollo de la genómica se ha disparado el número de secuencias disponibles en las bases de datos y el número de genomas completos disponibles (781 genomas de bacterias, 56 de archaea, 19 de eucariotas y 2524 virus según la versión de 2008 del Integrated Microbial Genome System (Markowitz *et al.*, 2008), perteneciendo un gran número de estos genomas a agentes patógenos. En la actualidad, además, la secuenciación completa de algunos virus se ha convertido en una herramienta rutinaria para su estudio.

Esta gran cantidad de secuencias procedentes de organismos patógenos, junto con la facilidad actual para acceder a técnicas de secuenciación, se han convertido en poderosas herramientas en salud pública dado que permiten trazar el origen y la evolución de epidemias o brotes causados por enfermedades infecciosas mediante la realización de estudios

filogenéticos (Bracho *et al.*, 2006; Bracho *et al.*, 2005; Esteban *et al.*, 1996). Esta abundancia de secuencias también ha servido como sustrato para una mejor comprensión de los procesos evolutivos de los agentes patógenos y sirve como base para el desarrollo de nuevas técnicas basadas en el análisis de ácidos nucleicos (nucleic acid testing, NAT).

En los últimos años se ha asistido al desarrollo de las denominadas aplicaciones post-genómicas. Las aplicaciones post-genómicas tales como la genómica comparativa, la genómica funcional, la genómica individual o la proteómica tienen cada vez una mayor aplicación en los entornos clínicos y se está produciendo la incorporación gradual de sus conocimientos y desarrollos en los laboratorios de microbiología clínica y las aplicaciones en salud pública. La genómica comparativa esta siendo utilizada para el estudio y la identificación de factores de virulencia o patogenicidad (Rajashekara *et al.*, 2004; Beare *et al.*, 2009; Hotopp *et al.*, 2006; Oberste, 2008). La genómica funcional ha sido empleada en un gran numero de aplicaciones que van desde el estudio de las alteraciones provocadas durante las infecciones a la búsqueda de nuevas dianas terapéuticas o la identificación de marcadores pronósticos o diagnósticos (de Lang A. *et al.*, 2007; Freiberg and Brotz-Oesterhelt, 2005; Rajashekara *et al.*, 2006). La genómica individual ha sido empleada en estudios epidemiológicos y evolutivos (Filliol *et al.*, 2006; Garcia Pelayo *et al.*, 2009; Touchman *et al.*, 2007). Finalmente las aplicaciones proteómicas se han usado para el desarrollo de vacunas o la identificación de marcadores (Neuman, V and Robinson, 2004; Steel *et al.*, 2005; Titball and Petrosino, 2007).

El desarrollo y aplicación de todas estas nuevas técnicas ha sido posible gracias al soporte suministrado por parte de la bioinformática para el almacenamiento y procesamiento de toda esta información. La maduración de la bioinformática como disciplina científica con un cuerpo de conocimiento propio se ha completado gracias a los nuevos retos planteados como consecuencia del desarrollo de la genómica y la conclusión de los diversos proyectos genoma. De esta manera las herramientas diseñadas para el análisis y comparación de secuencias son una herramienta fundamental para el desarrollo de las técnicas genómicas anteriormente comentadas. También hay que destacar el desarrollo de bases de datos con herramientas “*ad hoc*” para el análisis específico de algunos agentes de interés en salud pública como puede ser el VIH (Virus de la Inmunodeficiencia Humana) en las bases de datos mantenidas por el Laboratorio Nacional de Los Alamos en EE.UU.: (<http://www.hiv.lanl.gov>) o el virus de la gripe en el NCBI (National Center for Biotechnology Information, EE.UU.) (Bao *et al.*, 2008) proporcionando también herramientas para su utilización en las aplicaciones de inmunodiagnóstico.

Esta colaboración entre lo que se conoce como ciencia básica y ciencia aplicada se puede considerar como un ejemplo de transferencia de conocimientos y aplicaciones que ha sido fructífera en el pasado y que en el futuro ha de continuar proporcionando nuevas herramientas y conocimientos para el desarrollo de nuevos, más precisos y más sensibles sistemas de diagnóstico, tratamiento y control de las enfermedades infecciosas, entroncando con la denominada investigación biomédica traslacional.

1.2 Problemas de interés en Salud Pública susceptibles de ser analizados mediante técnicas moleculares

La utilización de técnicas moleculares se ha convertido en una práctica rutinaria en los laboratorios de salud pública y microbiología clínica. Existe un gran número de áreas y patologías en las que estas técnicas tienen en la actualidad una importancia capital con finalidad diagnóstica. A continuación se van a presentar tres problemas de salud pública seleccionados como modelos en los que es posible aplicar las técnicas moleculares.

1.2.1 Detección de agentes de potencial uso en bioterrorismo

La utilización con fines militares de los agentes infecciosos se remonta a la antigüedad, cuando ya se emplearon cadáveres de animales para contaminar pozos y fuentes de agua o se lanzaron cadáveres en ciudades sitiadas con el fin de propagar epidemias (Tucker, 1999). Esto sucedió en 1346 durante el sitio por los tártaros de la ciudad genovesa de Caffa quienes arrojaron al interior de la ciudad los cadáveres de víctimas de peste bubónica (Derbes, 1966), convirtiéndose de esta manera en el origen de la epidemia de peste. Esta infección fue provocada por la bacteria *Yersinia pestis*, que asoló Europa en el siglo XIV y es uno de los primeros ataques “biológicos” de los que se tiene constancia. Posteriormente se utilizaron mantas contaminadas con el virus de la viruela durante la guerra francobritánica en Norteamérica en el siglo XVIII. Más recientemente, en el siglo XX, durante la primera guerra mundial, el ejército alemán empleó las bacterias causantes del ántrax y el muermo (*Bacillus anthracis* y *Burkholderia mallei*) para infectar los animales de tiro de los aliados con la esperanza de causar problemas de abastecimiento y fue acusado de haber intentado propagar la peste en la ciudad rusa de San Petersburgo. A raíz de todos estos acontecimientos en el año 1925 se firmó el acuerdo conocido como Convención de Ginebra por el cual se prohibía la utilización durante la guerra de “gases asfixiantes, venenosos o de otro tipo y de métodos de guerra bacteriológica”. A pesar de la existencia de este tratado continuaron las investigaciones para el desarrollo de agentes de guerra biológica. Las tropas japonesas fueron acusadas durante la Segunda Guerra Mundial de haber atacado China, Mongolia y la Unión Soviética con peste así como de haber utilizado a prisioneros de guerra como sujetos de investigación con este tipo de agentes. En el año 1972 se firmó la Convención de Armas Biológicas, que entró en vigor en el año 1975, por el cual las naciones signatarias se comprometían a “...no desarrollar, almacenar o adquirir agentes biológicos o toxinas con finalidades distintas a las profilácticas, protectoras o pacíficas...”.

A pesar de todas estas medidas prohibitivas y de que el foco de interés ha estado siempre en el posible uso militar de estos agentes, la relativa simplicidad para su obtención, manejo y diseminación ha provocado que la utilización de agentes infecciosos se haya convertido en un área de interés para grupos terroristas. Hasta el año 1984 no se consideró el bioterrorismo como una amenaza seria. En ese año los seguidores de Bagwan Shree Rajneesh provocaron un brote (con mas de 750 casos) de *Salmonella typhimorium* contaminando las ensaladas de varios restaurantes en Oregon, EE.UU. Posteriormente en el año 1995 la secta

japonesa “Aum Shinrikyo” provocó un ataque con gas sarin en el metro de Tokio. Tras la detención de sus miembros se descubrió que además de este ataque con un gas nervioso también habían intentado varios ataques empleando agentes bacteriológicos entre los años 1990 y 1995 (Olson, 1999). Otros incidentes relacionados con bioterrorismo fueron la adquisición a través del ATCC (American Type Culture Collection) de 3 viales de *Yersinia pestis* por parte de un supremacista blanco de Ohio, EE.UU. en 1995 o el más reciente y mediático ataque mediante la dispersión intencionada de esporas de ántrax en el año 2001 en Estados Unidos por correo., que llegó a causar varias muertes.

Todos estos sucesos han provocado un interés creciente por parte de la salud pública, lo que ha movilizado recursos para coordinar y preparar las respuestas antes este tipo de amenazas mediante programas de vigilancia epidemiológica y respuesta frente a enfermedades infecciosas emergentes. Una de las áreas clave en estos programas es la relacionada con la disponibilidad de sistemas de detección y diagnóstico de estos agentes.

La asignación de un patógeno a la categoría de agentes susceptibles de ser empleados como arma bacteriológica o en bioterrorismo está relacionada con parámetros tales como la capacidad de transmisión del agente, la morbilidad y/o la mortalidad causada por el patógeno en cuestión. En la actualidad existen diferentes clasificaciones que catalogan distintos agentes en categorías (ver tabla 1) en función de los parámetros anteriormente descritos, siendo la clasificación desarrollada por el CDC (Centers for Disease Control and Prevention, www.cdc.gov) estadounidense la más frecuentemente empleada. En esta clasificación existen tres categorías:

- En la categoría A se recogen agentes que pueden ser fácilmente diseminados o transmitidos de persona a persona, tienen una elevada tasa de mortalidad con un elevado impacto potencial en salud pública, pueden causar pánico social y requieren una atención especial por parte de la salud pública.
- La categoría B contiene agentes con una capacidad de diseminación relativamente alta, causan una elevada morbilidad y baja mortalidad.
- Por último en la categoría C estarían encuadrados patógenos emergentes que son accesibles, fáciles de producir y diseminar y potencialmente causantes de alta morbilidad y mortalidad

En caso de producirse un ataque biológico se ha de contar con los medios para proporcionar una rápida respuesta que permita la interrupción de la propagación del agente y la extensión de las infecciones. Para ello es necesario contar con herramientas que permitan la detección temprana y precisa del agente o agentes implicados. Las técnicas moleculares surgen en este contexto como herramientas fundamentales para el desarrollo de ensayos diagnósticos que permitan la detección de estos agentes, siendo las reacciones de PCR las principales técnicas dirigidas hacia una identificación temprana debido a que otras técnicas

como el cultivo de los agentes o los ensayos serológicos para detectar la respuesta inmune de posibles afectados necesitan más tiempo para su realización (Buitrago Serna *et al.*, 2007).

Agente	Enfermedad	Categoría
<i>Bacillus anthracis</i>	Antrax	A
<i>Yersinia pestis</i>	Peste	A
<i>Francisella tularensis</i>	Tularemia	A
<i>Variola major</i>	Viruela	A
<i>Filovirus y Arenavirus</i>	Fiebres Hemorrágicas	A
<i>Brucella melitensis</i>	Brucelosis	B
<i>Burkholderia mallei</i>	Muermo	B
<i>Burkholderia pseudomallei</i>	Melioidosis	B
<i>Coxiella burnetii</i>	Fiebre Q	B
<i>Chlamydia psittaci</i>	Psitacosis	B
<i>Rickettsia spp</i>	Fiebre tifoidea	B
<i>Alphavirus</i>	Encefalitis	B
<i>Hantavirus</i>	Fiebre hemorrágica síndrome pulmonar	C
<i>Flavivirus</i>	Fiebre hemorrágica o encefalitis	C

Tabla 1. Agentes de potencial uso en bioterrorismo y su clasificación según las categorías del CDC

1.2.2 Identificación de agentes causantes de patologías respiratorias - Adenovirus

Los adenovirus humanos son virus icosaédricos sin envuelta. Su genoma está constituido por ADN de doble cadena y su tamaño oscila entre las 35 y las 38 Kb. La cápside de estos virus está constituida por grupos de tres proteínas diferentes, 12 proteínas fibrilares de anclaje, 12 proteínas penton relacionadas con el reconocimiento de los receptores celulares y 240 proteínas hexon que constituyen el principal componente de la cápside viral.

En la actualidad los adenovirus humanos representan una gran familia que comprende 52 serotipos diferentes que pueden ser subdivididos a su vez en 6 grupos (A a F) de acuerdo con diferentes propiedades biológicas tales como la oncogenicidad, la similitud nucleotídica, la hemaglutinación y características morfológicas relacionadas con las proteínas fibrilares.

La infección por adenovirus puede causar una gran variedad de enfermedades que pueden afectar a los tractos gastrointestinal, respiratorio y ocular (Potter, 1967; van, V, 1963; Laibson, 1975). Se ha observado que diferentes serotipos presentan un diferente tropismo para la infección de los diferentes órganos provocando diferentes patologías, siendo los adenovirus pertenecientes a los serotipos 1, 2, 3, 4, 5, 7 los que se identifican mayoritariamente en pacientes que sufren infecciones respiratorias agudas (Schmitz *et al.*, 1983; van, V *et al.*, 1969; Ryan *et al.*, 2002). Los adenovirus provocan frecuentes epidemias en niños (especialmente entre los recién nacidos (Abzug *et al.*, 1990; Elnifro *et al.*, 2005)), o pacientes con algún tipo de inmunodeficiencia (Pham *et al.*, 2003; Ebner *et al.*, 2006; Echavarria, 2008). En el caso de las infecciones por adenovirus en pacientes inmunodeficientes por causas tales como los trasplantes se pueden alcanzar tasas de mortalidad de hasta el 60% (Hierholzer, 1992) mientras que en las infecciones infantiles por adenovirus con tropismo respiratorio puede causar bronquiolitis y neumonía que, en algunos, casos pueden resultar fatales (Abzug and Levin, 1991; Klinger *et al.*, 1998).

La infección por adenovirus es una importante causa de enfermedades del tracto respiratorio siendo responsable de entre el 5 y 10% de las infecciones del tracto respiratorio inferior en niños. Estos virus se extienden rápidamente en ambientes cerrados y comunidades masificadas lo que favorece su aparición en forma de brotes epidémicos (van, V and DIJKMAN, 1962; van, V et al., 1969; Sanchez et al., 2001) pudiendo darse también de forma endémica. Algunos autores han apreciado diferencias en la forma de aparición de la enfermedad asociadas con el serotipo de los virus causantes de la infección asociando los serotipos 3 y 7 a brotes epidémicos (Ryan *et al.*, 2002) mientras que el resto de adenovirus con tropismo respiratorio están mayoritariamente asociados a infecciones esporádicas. Además de este diferente comportamiento epidemiológico, existen también evidencias que sugieren diferencias en la severidad de la morbilidad asociadas a los diferentes genotipos de los adenovirus (Gray *et al.*, 2005). En Suramérica se han asociado cepas de adenovirus 7h con una mayor severidad de la patología respiratoria, lo que provoca la necesidad de una mayor atención hospitalaria de los mismos (Carballal *et al.*, 2002; Larranaga *et al.*, 2000). Por estos motivos incluso existen propuestas para considerar a los adenovirus como patógenos emergentes (Gray, 2006).

En muchos casos, la infección por adenovirus resulta indistinguible de otro tipo de infecciones respiratorias causadas por virus o bacterias siendo por tanto necesaria la utilización de métodos y análisis microbiológicos para la confirmación de la infección por adenovirus.

Actualmente existe cierta controversia respecto a los estudios de respuesta frente a terapias empleando antivirales en los que se ha observado, "*in vitro*", sensibilidad dependiente del serotipo (Morfin *et al.*, 2009; Morfin *et al.*, 2005), por lo tanto, se hace necesario disponer de herramientas que permitan identificar no solo que se trata de una infección por adenovirus si no también el subtipo correspondiente. La disponibilidad de métodos moleculares de diagnóstico y tipificación rápidos ha de permitir la selección más adecuada de los tratamientos, cuando existan, evitando la utilización de tratamientos inadecuados o inefectivos. Estas metodologías han de permitir también identificar el origen de la infección, nosocomial, comunitaria, reactivación de una infección previa, lo que se traduciría en mejoras en el control de la transmisión de la infección, y redundaría en mejoras en la salud pública.

Las técnicas clásicas para la detección de los adenovirus han ido evolucionando con el tiempo hacia técnicas moleculares. Las técnicas clásicas empleadas para la detección y serotipado de la infección por adenovirus eran lentas y laboriosas e incluían abordajes tales como el cultivo celular (Belsy *et al.*, 2009), análisis de efectos citopáticos, ensayos de inmunofluorescencia (Gardner *et al.*, 1972), microscopia electrónica, hemaglutinación, ensayos de microneutralización (Malasig *et al.*, 2001). El principal inconveniente de estas técnicas radica en su lentitud a la hora de ofrecer resultados dado que estos pueden tardar entre 3 días y tres semanas dependiendo de las titulaciones virales presentes en las muestras (Echavarria *et al.*, 1998a), esta tardanza en ofrecer los resultados podría desembocar en su inutilidad para proponer tratamientos o controlar un brote epidémico. Como alternativa a estas técnicas se desarrollaron diferentes aproximaciones dirigidas a la detección directa del ácido nucleico del

virus mediante el desarrollo de técnicas de PCR. De este modo se han puesto a punto reacciones de PCR para el diagnóstico de estos virus (Avellon *et al.*, 2001; Casas *et al.*, 2005; Echavarria *et al.*, 2001; Echavarria *et al.*, 1998b; Faix *et al.*, 2004; Heim *et al.*, 2003). A partir de estas técnicas de PCR se han elaborado nuevos métodos en las que se han combinado PCR y secuenciación (Sarantis *et al.*, 2004; Takeuchi *et al.*, 1999), PCR y enzimas de restricción para el genotipado (Allard *et al.*, 2001) o incluso micromatrices (Lin *et al.*, 2004; Muller *et al.*, 2009).

1.2.3 Infección por VIH-1 y subtipos

El SIDA (Síndrome de Inmunodeficiencia Adquirida) es un grave problema de salud pública a nivel mundial. Esta enfermedad, una de las más graves pandemias del siglo XX, está provocada por la infección por el virus de la inmunodeficiencia humana (VIH) (Barre-Sinoussi *et al.*, 1983). El VIH es un retrovirus del género lentivirus. Se trata de un virus con envuelta en el que están expuestas las proteínas gp41 y gp120. En su interior la cápside formada por la proteína p24 protege las dos copias del RNA viral, y las enzimas que necesita el virus una vez penetra en la célula (transcriptasa inversa, proteasa, ribonucleasa e integrasa). El genoma viral tiene un tamaño de 9.181 nt con 9 genes que codifican 9 proteínas.

Una de las principales características de este virus es su gran heterogeneidad genética. Esta variabilidad está provocada por numerosos factores tales como la ausencia de un sistema de corrección de errores en su retrotranscriptasa, su elevada tasa de replicación y la presión del sistema inmune. Durante los últimos años se ha asistido a la diversificación del virus, lo que ha provocado problemas para el diagnóstico, la monitorización y el seguimiento de las terapias y es una de las principales causas para el fracaso de los intentos de desarrollo de vacunas. Como consecuencia de todas estas razones el estudio epidemiológico y filogenético del virus para su clasificación ha constituido un reto muy importante. En la actualidad se distinguen tres grupos dentro del VIH-1, M (main), O (outlier) y N (Non-M Non-O). El grupo M, responsable de la pandemia, comenzó su extensión en humanos hace aproximadamente unos 70 años. En este tiempo el virus se ha diversificado generando nueve subtipos (A-D, F-H, J y K) que pueden ser identificados mediante el análisis filogenético de los genes *env* y *gag* (Robertson *et al.*, 2000). Así mismo mediante estos análisis filogenéticos se han podido identificar procesos de recombinación entre diferentes subtipos. El primer caso de recombinación se observó cuando se analizó una cepa epidémica en Tailandia que originalmente fue considerada como el “subtipo E”. El análisis filogenético del gen *env* se agrupaba sólo de manera específica, mientras que el análisis del gen *gag* mostraba un agrupamiento con el subtipo A. A la vista de estos resultados se determinó que el “subtipo E” realmente era una forma recombinante entre un virus de subtipo A en la región *gag* y otro subtipo en la región *env*. En la actualidad se han identificado muchas formas recombinantes circulantes (CRFs) constituidas por combinaciones generadas por la coinfección con dos subtipos diferentes de virus y que han provocado sus propias epidemias (Fleury *et al.*, 2006; Holguin *et al.*, 2005; Holguin *et al.*, 2002; Holguin *et al.*, 2000; Papa *et al.*, 2002; Perez-Alvarez *et al.*, 2003).

Los diferentes subtipos del VIH, presentan diferentes características biológicas que modifican su interacción con el hospedador, lo que puede afectar a aspectos claves de la infección como la transmisión, el progreso de la enfermedad o la sensibilidad frente a los diferentes tratamientos disponibles (Dittmar *et al.*, 1997; Pope *et al.*, 1997; Dittmar *et al.*, 1997; Julg and Goebel, 2005; Kondo *et al.*, 2005; Neilson *et al.*, 1999; Renjifo *et al.*, 2001; Vergne *et al.*, 2003; Walter *et al.*, 2009).

La gran variabilidad genética del VIH supone un reto para el desarrollo de técnicas diagnósticas moleculares o serológicas que puede afectar tanto a la especificidad como a la sensibilidad de los ensayos. Un gran número de ensayos diagnósticos se basan en la detección de anticuerpos frente al VIH y la variabilidad del virus. La mayor dispersión del tipo M y el del subtipo B en los países occidentales provocó que inicialmente los sistemas estuviesen enfocados hacia la detección de estos virus, pero en la actualidad los sistemas son capaces de detectar todos los subtipos del grupo M y las formas recombinantes prevalentes. Además de las técnicas serológicas también se han desarrollado sistemas moleculares basados en la utilización de reacciones de PCR en tiempo real (Gueudin *et al.*, 2007; Stevens *et al.*, 2008).

Otro de los aspectos clave de la variabilidad genética viral es la posibilidad que ésta le otorga para eludir la respuesta inmune de los pacientes infectados. La gran variabilidad intra e inter subtipo de las proteínas de la envuelta (gp120 y gp41) hace que los esfuerzos realizados hasta el momento para el desarrollo de una vacuna preventiva hayan fracasado. El papel de la respuesta humoral frente a la infección por VIH es complejo. La infección por VIH induce pocos anticuerpos neutralizantes. Aparentemente los sitios de unión de los anticuerpos frente a epítomos conservados sufren cambios conformacionales que provocan impedimentos estéricos que reducen el reconocimiento efectivo de estos sitios por parte de los anticuerpos. Por estos motivos se está trabajando activamente en el estudio de la respuesta inmune humoral y celular con la esperanza de poder encontrar los epítomos adecuados para el desarrollo de una vacuna eficaz.

1.3 Micromatrices de material biológico

Las micromatrices de material biológico también conocidas como microarrays o biochips, constituyen uno de los resultados más significativos del desarrollo de las nuevas tecnologías para la obtención masiva de información biológica durante el desarrollo de la denominada era genómica.

1.3.1 Antecedentes

El desarrollo de esta tecnología se basa en la capacidad de inmovilizar sobre una superficie sólida una elevada densidad de material biológico (sondas), siguiendo un patrón regular y conocido (estructura en forma de matriz, que les da nombre). Esta elevada densidad sólo se puede lograr mediante la aplicación de técnicas procedentes de la miniaturización. Rápidamente y desde los comienzos de su desarrollo se mostraron como unas poderosas herramientas para la realización de ensayos masivos en los cuales se desean analizar

simultáneamente decenas de miles de genes (por ejemplo para analizar cuales se están expresando en un momento dado o analizar sus secuencias).

La idea de inmovilizar material biológico para la realización de ensayos se comenzó a desarrollar durante el siglo pasado, en los años 60. Sin embargo, el desarrollo de las micromatrices no fue posible hasta que se desarrollaron técnicas que permitieron la miniaturización y la síntesis “in situ” de material biológico. Gracias a la aparición de estas técnicas, algunas procedentes de la microelectrónica, a comienzos de los años 90 aparecieron las primeras publicaciones en las que se citaba esta tecnología (Maskos and Southern, 1992; Southern *et al.*, 1992). En la empresa Affymax, que posteriormente daría lugar a la actual Affymetrix, se ensayaron técnicas fotolitográficas utilizadas por la industria para la fabricación de los chips electrónicos para la síntesis de biopolímeros. Los ensayos iniciales estuvieron dirigidos inicialmente hacia la síntesis de péptidos sobre una superficie sólida pero posteriormente derivaron en la síntesis de oligonucleótidos (Fodor *et al.*, 1991). La utilización de esta técnica fotolitográfica, que fusionó el mundo de los chips electrónicos con la biología, es la que dio origen al término “biochip” o “GeneChip” con los cuales también se conoce a estos dispositivos. El término “biochip” ha llevado a la confusión en ocasiones ya que en otros contextos el término biochip ha sido empleado para referirse a otras tecnologías en las que interaccionan la biología y la tecnología electrónica. Algunos ejemplos de estas interacciones y en las que se pueden dar confusiones son la computación con ADN, el desarrollo de procesadores neuronales y los implantes de biodispositivos.

En la actualidad, existe una gran diversidad de tecnologías que permiten la fabricación de este tipo de dispositivos. Esta diversificación de técnicas permitió el desarrollo de robots capaces de inmovilizar material biológico presintetizado con el grado de miniaturización y precisión necesario para alcanzar las elevadas densidades de integración de material biológico requeridas en la fabricación de las micromatrices, pero sin alcanzar la densidad que se puede lograr mediante la utilización de los métodos fotolitográficos. La posibilidad de emplear estos robots, conocidos como arrayers o spotters, para la fabricación permitió un abaratamiento en los costes de fabricación y facilitó el acceso de los laboratorios a la tecnología, que inicialmente estaba limitado por los elevados costes de los procesos de fabricación por fotolitografía (Schena *et al.*, 1995). Existen diferentes técnicas como la fotolitografía o técnicas de inyección que permiten la fabricación directa de sondas de ácidos nucleicos sobre las superficies de la micromatriz mediante procesos de síntesis “in situ”. Este tipo de técnicas presenta algunas limitaciones en el material inmovilizado sobre la superficie dado que solo puede ser un ácido nucleico y con un tamaño inferior a 70 bases. A diferencia de lo que sucede con las técnicas de síntesis “in situ”, los arrayers, con la excepción de los que emplean técnicas piezoeléctricas de inyección, solo son capaces de inmovilizar material biológico que haya sido presintetizado. La gradual expansión de la técnica motivada por la llegada y la utilización de los arrayers junto con la capacidad de estos para depositar cualquier material que haya sido sintetizado u obtenido previamente, permitió a su vez el desarrollo de nuevos campos de aplicación y la superación

de algunas limitaciones de los procesos fotolitográficos, como la inmovilización de proteínas (MacBeath and Schreiber, 2000; Haab *et al.*, 2001; MacBeath, 2002).

1.3.2 Metodología de trabajo con micromatrices

En la realización de experimentos con se pueden diferenciar tres grandes fases: diseño y fabricación, preparación de las muestras y realización del ensayo y por último generación de resultados y análisis de los datos (ver figura 1).

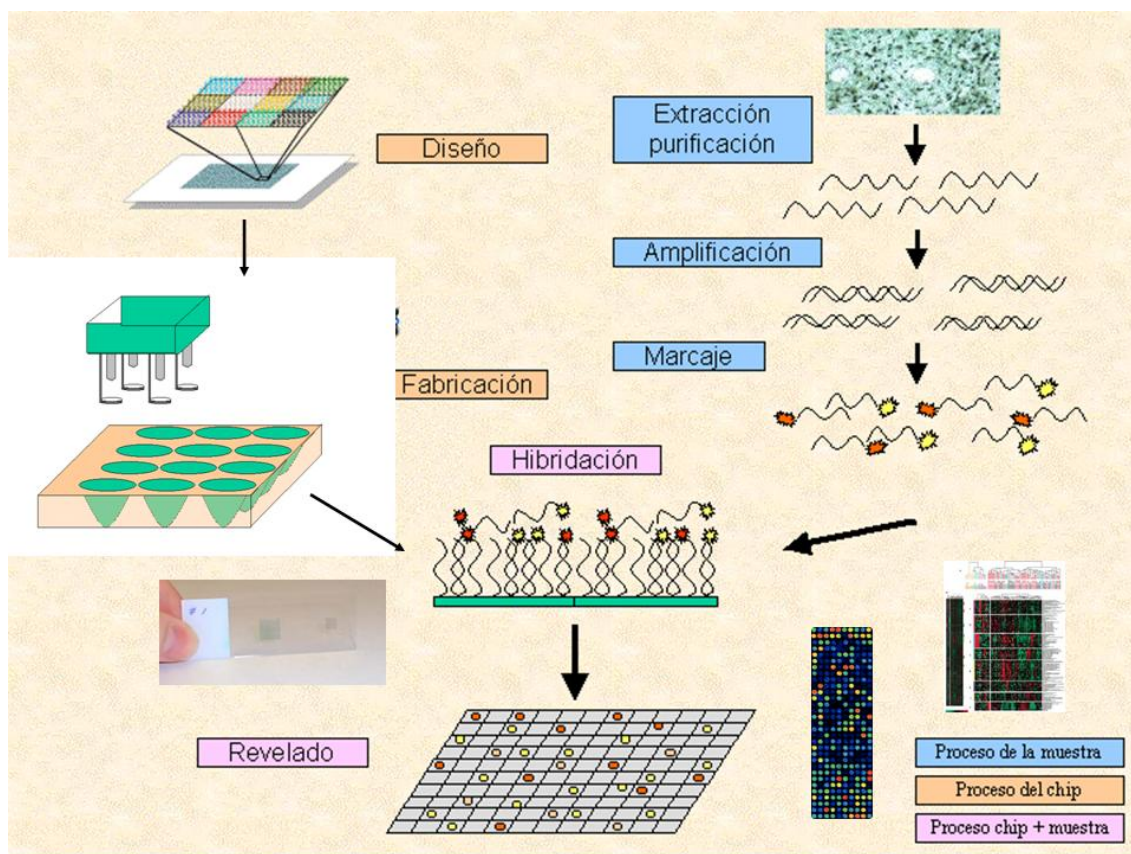


Figura1. Representación de la metodología a seguir durante un ensayo con micromatrices de DNA. Se pueden apreciar las fases de diseño y fabricación del dispositivo, tratamiento de las muestras, la reacción de afinidad (al tratarse de un ejemplo con ácidos nucleicos se realiza una hibridación), el revelado y el posterior análisis bioinformático de los datos.

La etapa de diseño y fabricación de las micromatrices está influenciada por el tipo de ensayo a realizar y por el conocimiento del sistema estudiado. En esta fase, se han de diseñar las sondas que se van a inmovilizar sobre la superficie de la micromatriz, se decide la densidad de integración y posteriormente se procede a la fabricación de las micromatrices empleando alguna de las técnicas de síntesis “in situ” o mediante la inmovilización de material presintetizado. Cuando se trabaja con micromatrices comerciales, esta fase es obviada. Es una etapa en la que se requiere de un gran soporte bioinformático para la gestión de equipos y de la información relacionada. En la sección dedicada a la informática de las micromatrices se detalla la aportación bioinformática en esta fase.

En la fase de preparación de muestras se realizan todos los procesos necesarios para poder realizar la posterior reacción de afinidad. En un esquema clásico y general del desarrollo

de experimentos con micromatrices, estos procesos consisten en la extracción y purificación del material de interés (ARN, ADN, proteínas, etc...), amplificación del mismo en los casos en los que sea necesario y por último la reacción de marcaje en la cual se produce la incorporación de los marcadores fluorescentes que permitan la posterior detección de las reacciones positivas. Este esquema general está condicionado por el tipo de ensayo a realizar y por el tipo de muestra, (por ejemplo en ensayos de extensión de cebador no se marcan las muestras y en ensayos con proteínas no se produce amplificación de las muestras).

Finalmente y una vez están listas las muestras, se procede a la realización de los ensayos sobre la superficie de las micromatrices. Inicialmente, la tecnología se centró en el desarrollo de ensayos de hibridación entre las sondas inmovilizadas sobre la superficie de las micromatrices y las muestras a analizar (Maskos and Southern, 1993; Williams *et al.*, 1994). Dependiendo del tipo de estudio estas hibridaciones eran sencillas, en las que se emplea una única muestra (Guo *et al.*, 1994), o competitivas en las que se hibridan simultáneamente dos muestras marcadas (DeRisi *et al.*, 1996; Shalon *et al.*, 1996) con dos fluoróforos diferentes. Con posterioridad al desarrollo de los ensayos de hibridación se desarrollaron nuevas aplicaciones en las que se incluían reacciones enzimáticas, usando ligasas o polimerasas (Pastinen *et al.*, 1997; Kurg *et al.*, 2000), sobre la superficie de las micromatrices, reacciones antígeno-anticuerpo (Haab *et al.*, 2001) y otro tipo de interacciones (Falsey *et al.*, 2001).

Una vez realizada la reacción sobre la superficie de la micromatriz es necesario realizar unos lavados para eliminar el material que no haya reaccionado y las reacciones inespecíficas que se hayan podido producir, dejando listas las micromatrices para al proceso de escaneado. Durante este escaneado se detecta la fluorescencia procedente de las posiciones en las que se ha producido una reacción positiva y se genera una imagen que va a ser el resultado del ensayo realizado.

La última fase de la metodología consiste en el análisis de los resultados. Esta fase es por lo general el cuello de botella en esta metodología debido a la gran cantidad de información que se debe procesar y a la complejidad de los análisis que suelen ser necesarios. Es una fase fundamentalmente bioinformática en la que se incluye el análisis y cuantificación de las imágenes y el posterior análisis numérico de los datos cuantificados.

1.3.3 Bioinformática asociada a las micromatrices

Los ensayos con micromatrices son ensayos en los que hay que gestionar una gran cantidad de información procedente de los experimentos así como la introducida para la anotación y descripción de los mismos. La bioinformática está presente en casi todos los procesos del trabajo con las micromatrices, desde los procesos de diseño de sondas (Plasterer, 1997; Rimour *et al.*, 2005; Wernersson and Nielsen, 2005; Li and Ying, 2006; Macula *et al.*, 2008) y la gestión de los equipos de fabricación hasta la fase final de análisis y presentación de los resultados (Tarraga *et al.*, 2008). Incluso en la etapa de preparación de muestras tiene cabida la bioinformática mediante los sistemas de gestión de información de

laboratorio (LIMS) (Maurer *et al.*, 2005), en los que se anotan todos los procesos que sufre la muestra desde su generación y/o recepción hasta que está lista para ser empleada sobre la superficie de la micromatriz. Todo este conjunto de actividades ha generado un conjunto de herramientas, aplicaciones y conocimiento que se conocen popularmente como “Array Informatics” o “Microarray Informatics”. La bioinformática enfocada en los estudios de las micromatrices tiene cuatro grandes focos de atención:

1. Diseño y anotación de sondas. El diseño de sondas consiste en el conjunto de procedimientos por el cual se selecciona una determinada región de una secuencia para posteriormente inmovilizarla sobre la superficie de la micromatriz. El proceso de diseño de sondas está íntimamente ligado al tipo de plataforma empleado para el desarrollo de los experimentos, la finalidad y tipo del ensayo a realizar.
2. Análisis y cuantificación de las imágenes. En esta etapa se produce la transformación de la imagen generada durante el proceso de escaneado a datos numéricos. Las imágenes que se generan durante el escaneado son imágenes en una escala de grises con una profundidad de color de 16-bits (lo que se traduce en aproximadamente 65.500 tonos de gris diferentes) con unas resoluciones que oscilan entre los 2 y los 10 píxel por micrómetro que suelen ser almacenadas en formato “.tiff”. La “cuantificación” de las imágenes es el proceso por el cual se asignan valores numéricos de intensidad a los puntos en los que se ha inmovilizado el material biológico. Este proceso es un proceso dividido en tres grandes etapas: “gridding” o asignación del mapa de localización de los puntos de material biológico; segmentación de las imágenes en regiones de señal y de ruido de fondo; y atribución numérica de intensidad (Qin *et al.*, 2005).
3. Análisis de los datos. Los datos numéricos generados durante el proceso de cuantificación de las imágenes han de sufrir numerosas etapas de control de calidad y preprocesamiento antes de llegar a alcanzar la fase final en la que con unos datos depurados se realiza el proceso analítico destinado a dar respuesta a las hipótesis planteadas por el experimento. Dentro de las tareas que se desarrollan en el preprocesamiento de los datos deben destacarse el análisis y control de calidad de los datos con los que se va a trabajar, la identificación y eliminación mediante diferentes técnicas de filtrado de los valores anómalos y la normalización. La normalización es una etapa muy importante ya que se trata del proceso que va a permitir la comparación entre datos procedentes de diferentes muestras, diferentes micromatrices etc... Durante el proceso de normalización los datos con los que se va a trabajar únicamente son transformados sin producir alteraciones en su contenido para corregir las desviaciones que se puedan haber producido en los mismos.

En general los experimentos con micromatrices se pueden clasificar en función del tipo de estudio que se desee realizar en tres grandes categorías: Estudios de comparación de clases, de predicción de clases y de descubrimiento de clases.

Los estudios de comparación de clases: son aquellos en los que se conocen las diferentes categorías o clases que conforman el experimento y en los que se quiere averiguar cuáles son las características propias de cada uno de ellos. Un ejemplo clásico de este tipo de ensayos es la determinación de los genes que están activados o desactivados en los procesos tumorales en comparación con aquellos que lo están en condiciones normales en una muestra procedente de un individuo sano.

Los estudios de predicción de clases: tratan de generar “predictores” capaces de incluir una muestra perteneciente a una clase desconocida en una o más clases previamente identificadas. Como ejemplo de este tipo de ensayos se podría citar un estudio en el que tras analizar muestras sanas y enfermas se busca identificar un perfil genético capaz de predecir si una muestra desconocida va a pertenecer a un paciente sano o enfermo.

Los estudios de descubrimiento de clases: son aquellos en los que no se emplea ningún tipo de asunción “a priori” sobre las características de los elementos analizados y se trata de descubrir si entre ellos existe algún tipo de asociación o relación.

4. Gestión, almacenamiento y anotación de los experimentos. Uno de los aspectos claves durante todo el proceso de los ensayos con las micromatrices es el control de la calidad y la anotación de todos y cada uno de los procesos realizados durante los ensayos. La gran cantidad de información que se ha de manejar en todas las etapas desde el diseño de las sondas hasta el análisis final hace que este proceso de seguimiento y anotación resulte crucial. Un aspecto clave para lograr esta calidad ha sido el desarrollo de diferentes herramientas de estandarización entre las que cabe destacar un estándar que cubre la información mínima necesaria que se ha de suministrar al referirse a los ensayos con micromatrices (MIAME, Minimum Information About a Microarray Experiment) (Brazma *et al.*, 2001; Knudsen and Daston, 2005). Junto con este estándar se han desarrollado también otros estándares dirigidos a facilitar la anotación de los procedimientos empleando una ontología común (Whetzel *et al.*, 2006) y el intercambio de datos mediante un estándar basado en XML (Spellman *et al.*, 2002). En la actualidad y desde la adopción de los estándares propuesto por MGED, se ha logrado unificar en torno a estos estándares la información almacenada en las bases de datos y la forma de transmitirla. En la actualidad existen varias bases de datos públicas como GEO (Barrett *et al.*, 2009), ArrayExpress (Parkinson *et al.*, 2009) o SMD (Hubble *et al.*, 2009), en las que se pueden consultar una gran cantidad de ensayos y en las que es posible depositar los datos de nuevos experimentos.

1.3.4 Aplicaciones de las micromatrices

Las micromatrices son una de las herramientas más potentes de las que se dispone en los laboratorios actuales debido a su capacidad para la realización de un gran número de determinaciones en paralelo sobre una muestra. Desde sus comienzos estas técnicas se han

focalizado en el desarrollo de análisis genómicos, pero en la actualidad debido a los nuevos desarrollos se están aplicando en áreas como la proteómica o metabolómica.

Los estudios con micromatrices pueden ser analizados desde varias perspectivas, dependiendo del material inmovilizado (ADN, proteínas u otras moléculas), del tipo de material estudiado (ADN, ARN, proteínas u otros) o del tipo de determinación que se desee realizar (cualitativas o cuantitativas).

En los ensayos cualitativos, se analiza la presencia de uno o más analitos presentes en una muestra, en determinaciones de tipo “Sí/No”, sin tener en cuenta la cantidad presente de éste. Estas aproximaciones se emplean en el estudio del ADN, especialmente en estudios de resecuenciación o genotipado (Syvanen, 2005; Hao *et al.*, 2008; Garcia Pelayo *et al.*, 2009; Xing *et al.*, 2009), en un gran número de ensayos de proteínas (Hartmann *et al.*, 2009b; Wingren and Borrebaeck, 2008; O'Looney *et al.*, 2008; Chen *et al.*, 2003; Haab *et al.*, 2001) y es el tipo de aproximación seguida en el desarrollo de ensayos para la detección de agentes infecciosos. (Wilson *et al.*, 2002; Morales Espinosa *et al.*, 2003; Hartmann *et al.*, 2009a; Burgess *et al.*, 2008; Zhu *et al.*, 2006).

Los ensayos cuantitativos cubren los ensayos más frecuentes y extendidos de las micromatrices, constituyendo el cuerpo principal de las publicaciones con estas técnicas. En este tipo de ensayos se establece una relación entre la señal obtenida en la reacción de afinidad sobre la micromatriz y la cantidad de un determinado analito presente en las muestras. Las determinaciones cuantitativas siempre son relativas y comparativas, no pudiendo hacerse cuantificaciones directas en las que se determine la concentración exacta o el número de moléculas presentes en una muestra. Esta limitación proviene de la incapacidad para conocer con exactitud la cantidad de material que queda realmente inmovilizado en los puntos sobre la superficie del micromatriz. Dentro de los ensayos cuantitativos las aplicaciones para el estudio de los ARNm, en la expresión génica, son la principal área de estudio y aplicación de esta tecnología y han supuesto una de las bases para el desarrollo de la genómica funcional. (Mehra and Kaushal, 2009; Yin *et al.*, 2008; DeRisi *et al.*, 1996; Schena *et al.*, 1996; Schena *et al.*, 1995). Dentro de los ensayos cuantitativos el estudio cuantitativo del ADN mediante el análisis del número de copias de determinados genes o secuencias del genoma mediante técnicas de hibridación genómica comparativa (CGH), se está convirtiendo en una herramienta de uso frecuente

2. Objetivos

Las micromatrices o microarrays son un conjunto de técnicas que se han empleado principalmente para el análisis de la expresión génica y se han extendido en los últimos años. La utilización de estas técnicas en otras aplicaciones es todavía menos frecuente. La posibilidad de emplear estas técnicas para el estudio de la variabilidad de las secuencias de ácidos nucleicos y el desarrollo de nuevas aproximaciones en las que el material inmovilizado sobre la superficie o estudiado no es un ácido nucleico, puede suponer avances en el estudio y la investigación de los sistemas biológicos. En la presente tesis, se han empleado estos métodos en conjunción con la bioinformática para el estudio y desarrollo de nuevas técnicas de detección de agentes microbiológicos de interés en salud pública.

Los objetivos planteados en este trabajo son los que se detallan a continuación:

1. Estudiar la aplicación de las micromatrices de material biológico para la detección de patógenos humanos de interés en salud pública. Este objetivo general tiene a su vez los siguientes objetivos específicos:
 - a. Comparar y evaluar diferentes superficies para la inmovilización de material biológico durante la fabricación de micromatrices.
 - b. Establecer, analizar y validar un conjunto de sondas adecuadas para la detección de agentes susceptibles de ser empleados en bioterrorismo (de categoría A y B). Se distinguirá entre aproximaciones basadas en hibridación y en extensión de cebador.
 - c. Determinar un conjunto de sondas válidas para la detección de diferentes serotipos de adenovirus respiratorios para su posterior validación y análisis en muestras clínicas
2. Responder a las necesidades bioinformáticas para el diseño de sondas válidas para su utilización e inmovilización sobre las micromatrices, poniendo a disposición de la comunidad científica las herramientas desarrolladas.
3. Estudiar mediante micromatrices la respuesta inmune humoral frente a diferentes subtipos de VIH. Este objetivo general tiene a su vez los siguientes objetivos específicos:
 - a. Analizar y determinar la diferente reactividad de la respuesta humoral en la infección por diferentes subtipos de VIH.
 - b. Comparar los resultados obtenidos mediante técnicas ELISA y microarrays.
 - c. Proponer y estudiar nuevos abordajes para la determinación de perfiles de reactividad basados en la afinidad.

3. Materiales y métodos

3.1 Obtención y Tratamiento de las muestras

3.1.1 Muestras de agentes infecciosos de interés en Bioterrorismo

Para la validación de los métodos y dispositivos diseñados para la detección de agentes de interés en Bioterrorismo se emplearon muestras de DNA suministradas por los grupos colaboradores del Laboratorio de Espiroquetas y Patógenos Especiales del Centro Nacional de Microbiología (Instituto de Salud Carlos III) y del Laboratorio de Biología de la Fabrica Nacional de la Maraños (Mº de Defensa). Ver Tabla 2

Organismo	Tipo de muestra	Origen
<i>Bacillus anthracis</i> cepa Ames	DNA genómico	1
<i>Yersinia pestis</i> cepa 10362	DNA genómico	1
<i>Yersinia pestis</i> gen pla	Plásmido	1
<i>Francisella tularensis</i> , subsp. <i>tularensis</i> cepa B-38,	DNA genómico	1
<i>Francisella tularensis</i> subsp. <i>holarctica</i> cepa LVS	DNA genómico	1
<i>Francisella tularensis</i> subsp. <i>novicida</i> cepa Utah 112	DNA genómico	1
<i>Francisella tularensis</i> subsp. <i>philomiragia</i>	DNA genómico	1
<i>Burkholderia mallei</i> ,	DNA genómico	1
<i>Burkholderia pseudomallei</i> ,	DNA genómico	1
<i>Burkholderia thailandensis</i>	DNA genómico	1
<i>Burkholderia vietnamensis</i>	DNA genómico	1
<i>Brucella melitensis</i> subsp. <i>melitensis</i> cepa 16M	DNA genómico	1
<i>Brucella melitensis</i> subsp. <i>abortus</i> 86/8/59	DNA genómico	1
<i>Brucella melitensis</i> subsp. <i>suis</i> , cepa 686	DNA genómico	1
<i>Brucella melitensis</i> subsp. <i>canis</i>	DNA genómico	1
<i>Brucella melitensis</i> subsp. <i>pinnipeda</i> B2/94	DNA genómico	2
<i>Brucella melitensis</i> subsp. <i>melitensis</i> cepa Ether	DNA genómico	1
<i>Brucella melitensis</i> subsp. <i>abortus</i> cepa B19	DNA genómico	2
<i>Brucella melitensis</i> subsp. <i>suis</i> , cepa Thomson	DNA genómico	2

<i>Brucella melitensis</i> subsp. <i>ovis</i> cepa 331	DNA genómico	2
<i>Brucella melitensis</i> subsp. <i>ovis</i> cepa PA	DNA genómico	2
1 – Laboratorio de Espiroquetas y Patógenos especiales, Centro Nacional de Microbiología, Instituto de Salud Carlos III 2- Laboratorio de Biología, Fábrica Nacional de “La Marañosa”, Ministerio de Defensa		

Tabla 2 Muestras empleadas en los ensayos de detección de patógenos de interés en bioterrorismo, tipo de muestra proporcionada y origen de la muestra.

Para la validación de ensayos dirigidos a la detección del virus de la viruela (variola minor) fue necesaria la generación de un DNA quimérico artificial, que fue generado en nuestro laboratorio para ser empleado como muestra de validación. Con este objetivo se desarrolló una PCR para la extensión secuencial de oligonucleótidos artificiales.

Las reacciones de PCR para la extensión de los oligonucleótidos artificiales se realizaron en un volumen de 50µl con 10mM Tris-HCl, 50mM KCl, 4mM MgCl₂ y 200µM de cada uno de los deoxinucleósidos trifosfato (dNTPs) (GE Healthcare, Uppsala, Suecia) y 2 U de Taq DNA Polymerase (GE Healthcare, Uppsala, Suecia). Los oligonucleótidos artificiales (ver tabla 3) a extender en cada una de las rondas de PCR se emplearon en una concentración final de 0.4µM. Los ciclos de las reacciones de PCR incluyeron un ciclo de desnaturalización inicial de 3 min. a 95 °C seguidos de 40 ciclos de 30s a 95 °C, 30s a 50 °C y 30s a 72 °C y una etapa de elongación final de 3 minutos a 72 °C. Para la obtención de la quimera en una única etapa de PCR se sustituyó la mezcla que contenía el par de oligonucleótidos a elongar por una mezcla que contenía los 4 pares de oligonucleótidos más internos (S1, S2, S3, S4, As1, As2, As3 y As4) a una concentración final de 0.1µM y el par externo (oligonucleótidos S5 y As5) a una concentración final de 0.4µM, manteniéndose las mismas condiciones de ciclado en el termociclador.

Nombre	Secuencia 5'-3'
S5	TAT TAC TAT GAA TCA TAC AGA TTG CAA TCC CGT ATT TCG TGA GGA ATA CTT CTC T
S4	TAT TTC GTG AGG AAT ACT TCT CTG TCC TTA ATA AGG TAG CAA CTT C
S3	GTC CTT AAT AAG GTA GCA ACT TCA GGA TTT TTT ACA GGA GAA AAT AGA TAT CA
S2	TTT ACA GGA GAA AAT AGA TAT CAA AAT ATT TCA AAG GTG TG
S1	TAT CAA AAT ATT TCA AAG GTG TGT ACT TTA AAT TTT GAG ATT AAA TGT AAT AAC AAA GG
As1	TTC GTT AGC TGT TTG AAG GAA GAA CCT TTG TTA TTA CAT TTA ATC TC
As2	GAC ATC ATA CCG TCA TCA TTC TTT GCT TTC GTT AGC TGT TTG AAG G
As3	CGC TAG AGT TAC CGT CTC CGA ATG CGA CAT CAT ACC GTC ATC AT
As4	AGA TGT CGA CGC TAG ATA GAC AGT CAC CCG CTA GAG TTA CCG TCT C
As5	CAT TGG TAT TAC TAT ATA GTA TAT AGA TGT CGA CGC TAG ATA GAC

Tabla 3 Oligonucleótidos artificiales empleados para la generación de un control positivo para el virus de la viruela

3.1.2 Muestras de virus respiratorios

Los ensayos de detección de virus respiratorios emplearon muestras prototipo y muestras clínicas. Todas estas muestras fueron proporcionadas por el Laboratorio de Virus Respiratorios del Centro Nacional de Microbiología del Instituto de Salud Carlos III.

Los ensayos de sensibilidad emplearon controles basados en la utilización de 3 diluciones que contenían 10, 100 y 1000 copias del producto de amplificación del adenovirus 1 clonado. El producto de amplificación primario del adenovirus de serotipo 1 clonado en un vector pGMT-T ha sido descrito con anterioridad (Avellon *et al.*, 2001).

Se emplearon muestras de referencia procedentes de 49 adenovirus de diferentes serotipos, usándose los 1, 2, 3, 4, 5, 7, 14 y 21 como controles positivos y el resto de los serotipos (1-49 excepto los anteriormente referenciados) para los estudios de especificidad.

Entre estos controles de especificidad se incluyeron también productos de PCR clonados del virus respiratorio sincitial A y B y de cada tipo de virus influenza (A, B y C). Estos controles fueron previamente descritos en (Coiras *et al.*, 2003).

El DNA de los especímenes clínicos analizados fue extraído en el Laboratorio de Virus Respiratorios a partir de aspirados nasofaríngeos de niños que presentaban síndromes respiratorios agudos tales como bronquiolitis, neumonía o patologías similares a la infección gripal. Las muestras fueron analizadas empleando técnicas IFI (Inmunofluorescencia indirecta) y de PCR para la identificación de los principales virus respiratorios, incluyendo los adenovirus.

3.1.3 Muestras para el análisis de la respuesta inmune humoral

Se emplearon sueros de pacientes no infectados e infectados por VIH suministradas por el laboratorio de caracterización molecular del Servicio de Diagnóstico y Referencia de Retrovirus, del Centro Nacional de Microbiología, caracterizadas genéticamente en gp41 como de subtipo B, y sueros correspondientes a los años 1994 a 1997, del laboratorio de diagnóstico serológico del Servicio de Diagnóstico y Referencia de Retrovirus del Centro Nacional de Microbiología. Estas muestras se complementaron con muestras procedentes de un panel comercial internacional (BBI Boston Biomedica Inc., West Bridgewater, EE.UU.) . Tablas 4 y 5

Subtipo (Caracterizado en gp41)	Número de sueros analizados
A	3
B	14
C	3
D	3
E (CRF01_A/E)	3
F	3
G	2
Tipo O	2
Negativo	8

Tabla 4 Número de sueros analizados pertenecientes a cada subtipo en el estudio de la respuesta humoral frente a diferentes subtipos de VIH-1. En total en este análisis se emplearon 41 sueros

Muestra	Subtipo (Caracterizado en gp41)	Origen
ST-461	B	1
ST-711	B	1
ST-732	B	1
ST-453	B	1
ST-455	B	1
ST-661	B	1
ST-372	B	1
ST-440	B – Negativo en ELISA	1
ST-670	B – Negativo en ELISA	1
ST-794	B – Negativo en ELISA	1
ST-941	B – Negativo en ELISA	1
WW-301-17	B	2
WW-301-2	B	2
WW-301-32	B	2
ST-687	A	1
WW-301-19	A	2
WW-301-12	A	2
WW-301-23	D	2
WW-301-25	D	2
WW-301-26	D	2
WW-301-18	C	2
WW-301-28	D	2
WW-301-31	D	2
WW-301-3	G	2
WW-301-8	G	2
WW-301-38	E (CRF01_A/E)	2
WW-301-39	E (CRF01_A/E)	2
WW-301-40	E (CRF01_A/E)	2
WW-301-47	F	2
WW-301-48	F	2
WW-301-49	F	2
WW-301-1	Tipo O	2
1555/E/98	Tipo O	1
C(-)-1	Negativo	1
C(-)-11	Negativo	1
C(-)-12	Negativo	1
C(-)-15	Negativo	1
C(-)-16	Negativo	1
C(-)-2	Negativo	1
C(-)-5	Negativo	1
1- Laboratorio de caracterización molecular del Servicio de Diagnóstico y Referencia de Retrovirus, del Centro Nacional de Microbiología, Instituto de Salud Carlos III		
2- Boston Biomedica Inc		

Tabla 5. Muestras empleadas en los ensayos de análisis de la respuesta inmune humoral frente a diferentes subtipos de VIH-1

3.2 Reacciones de amplificación y marcaje

En la tabla 6 se recogen las referencias a las reacciones de amplificación empleadas para cada uno de los organismos de interés. Estas reacciones de amplificación fueron modificadas ligeramente para permitir el marcaje fluorescente de productos de amplificación mediante la incorporación de dCTP-Cy3 (GE Healthcare, Uppsala, Suecia) en una relación de 1:9 respecto al dCTP sin marcar.

Organismo	Fragmento amplificado	Referencia
F. tularensis	<i>16srna</i>	(Forsman <i>et al.</i> , 1994)
	<i>tul4</i>	(Long <i>et al.</i> , 1993; Sellek <i>et al.</i> , 2008)
Brucella	<i>bcs31</i>	Desarrollo propio
	<i>omp2</i>	Desarrollo propio
	<i>omp31</i>	Desarrollo propio
Burkholderia	<i>scu2</i>	(Thibault <i>et al.</i> , 2004)
	<i>orf11</i>	(Thibault <i>et al.</i> , 2004)
	<i>orf13</i>	(Thibault <i>et al.</i> , 2004)
	<i>lysE</i>	Desarrollo propio
	<i>gyrA</i>	Desarrollo propio
B. anthracis	<i>ba813</i>	(Ramisse <i>et al.</i> , 1996)
	<i>Lef</i>	(Ramisse <i>et al.</i> , 1996)
	<i>lef 3/4</i>	(Ramisse <i>et al.</i> , 1996)
	<i>Cya</i>	(Ramisse <i>et al.</i> , 1996)
	<i>Capbca</i>	(Ramisse <i>et al.</i> , 1996)
	<i>Pag</i>	(Ramisse <i>et al.</i> , 1996)
C. burnetii	<i>16srna</i>	Comunicación personal del Dr. P. Anda.
Y. pestis	<i>Pla</i>	(Engelthaler <i>et al.</i> , 1999)
Orthopoxvirus	<i>Crn</i>	(Sanchez-Seco <i>et al.</i> , 2006)

Tabla 6 con las reacciones de PCR empleadas en el diseño de una micromatriz para la identificación de agentes susceptibles de ser empleados en bioterrorismo.

En aquellos casos en los que se tuvieron que desarrollar nuevas reacciones de PCR para la amplificación de regiones de interés, las condiciones de PCR se recogen en las tablas 7 y 8.

	Condiciones de amplificación			
	<i>lysE</i>	<i>gyrA</i>	<i>bcs31</i> y <i>omp31</i>	<i>omp2</i>
Reactivo	Concentración final			
Primer Foward 1	0.5µM	0.5µM	0.5µM	0,8 µM
Primer Foward 2				0,8 µM
Primer Reverse	0.5 µM	0.5 µM	0.5 µM	0,8 µM
dNTPs	0,2 mM	0,2 mM	0,2 mM	0,2 mM
MgCl2	1.5 mM	2.5 mM	2 mM	2 mM
Taq	1 unit	1 unit	1 unit	1 unit
10Xbuffer	1x	1x	1x	1x
DNA				
H2O				
Vfinal	25	25	25	25

Tabla 7. Condiciones de las reacciones de PCR desarrolladas para la detección de miembros del género *Burkholderia* (reacciones de amplificación de *lysE* y *gyrA*) y *Brucella* (reacciones de amplificación de *bcs31*, *omp31* y *omp2*)

lysE	Condiciones de amplificación – Perfil térmico		
	gyrA	bcsp31 y omp31	omp2
95°C - 2 min.	95°C - 2 min.	95°C - 2 min.	95°C - 2 min.
40x (95°C – 30s; 50°C – 30s; 72°C – 45s)	40x (95°C – 30s; 50°C – 30s; 72°C – 45s)	40x (95°C – 30s; 50°C – 30s; 72°C – 45s)	3x (95°C – 30s; 68°C – 30s; 72°C – 45s)
72°C – 5 min.	72°C – 5 min.	72°C – 5 min.	2x (95°C – 30s; 65°C – 30s; 72°C – 45s)
			2x (95°C – 30s; 62°C – 30s; 72°C – 45s)
			2x (95°C – 30s; 59°C – 30s; 72°C – 45s)
			2x (95°C – 30s; 56°C – 30s; 72°C – 45s)
			30x (95°C – 30s; 53°C – 30s; 72°C 45s)
			72°C – 5 min.

Tabla 8. Condiciones de temperatura y ciclos de las reacciones de PCR desarrolladas para la detección del género *Burkholderia* (reacciones de amplificación de *lysE* y *gyrA*) y *Brucella* (reacciones de amplificación de *bcsp31*, *omp31* y *omp2*)

3.3 Diseño de cebadores y sondas

Para el diseño de los cebadores empleados en las reacciones de PCR desarrolladas en este trabajo se utilizó el programa de análisis de secuencias Oligo 6 (Molecular Biology Insights Inc., Cascade, CO, EE.UU.) Las secuencias para el diseño de las reacciones de PCR usadas para la identificación de las especies del género *Burkholderia* fueron seleccionadas a partir del análisis de las secuencias de los genomas de las diferentes especies de este género disponibles a través del Comprehensive Microbial Resource (CRM, <http://cmr.jcvi.org>) (Peterson et al., 2001a). En el caso de las reacciones para la amplificación de los genes *omp2*, *omp31* y *bcsp31* del género *Brucella* se recuperaron de Genbank las secuencias disponibles y tras realizarse un alineamiento múltiple usando el programa Clustal X (Thompson *et al.*, 2002) se procedió al diseño de los cebadores en las regiones conservadas de estos alineamientos.

El proceso bioinformático seguido para el diseño con las sondas empleadas en los ensayos de detección de los agentes infecciosos mediante micromatrices de ácidos nucleicos se realizó de manera que todas las sondas diseñadas compartieran unas características físico-químicas similares que permitan su empleo bajo las mismas condiciones de reacción.

3.4 Funcionalización de superficies y Fabricación de las micromatrices

Para el proceso de fabricación de las micromatrices se han empleado portaobjetos de vidrio funcionalizados en nuestro laboratorio. Las sondas se depositaron sobre la superficie de las micromatrices a una concentración de 40 μM en tampón carbonato pH=9 o 3xSSC, dependiendo del tipo de superficie, empleando un robot GMS-417 (Affymetrix, Santa Clara, CA, EE.UU.).

El tipo de unión de las sondas a la superficie de las micromatrices estuvo definido por el tipo de superficie generada durante el proceso de funcionalización. En el desarrollo de estos trabajos se analizó el comportamiento de cinco tipos de funcionalizaciones diferentes, Poly-L-lisina, aminosilano, aminoepoxisilano, epoxisilano e isotiocianato.

La funcionalización de los portaobjetos con Poly-L-Lisina se realizó siguiendo el protocolo descrito anteriormente por (Schena et al., 1995). Para la funcionalización con aminosilano, epoxisilano y aminoepoxisilano, se tomó como base el protocolo descrito en Guo et al., 1994 y se hicieron ligeras modificaciones para permitir la funcionalización con los otros grupos indicados anteriormente. El proceso de funcionalización con isotiocianato se basó en el descrito por Lindroos et al., 2001. En la tabla 9 se presentan brevemente los protocolos empleados en estos procesos de funcionalización e impresión.

Protocolos de funcionalización de portaobjetos de vidrio para la inmovilización de material biológico

Lavado y preparación de los portaobjetos con NaOH·EtOH

1. Lavado durante dos horas de los portaobjetos en una solución de hidróxido sódico 6.25M y etanol en una relación 4:7 v/v
2. 5x Aclarado en ddH₂O para retirar cualquier resto de la solución de lavado
3. Centrifugado 2 min. a 700rpm

Funcionalización con Aminosilano:

Se usan los portaobjetos lavados con NaOH·EtOH

1. 5 min. En una solución 1% de γ -aminopropiltrimetoxisilano en 95% acetona
2. 2x 2 min. En acetona
3. 2x 2 min. En etanol
4. Centrifugado 3 min. a 700rpm
5. Hornear a 85°C durante toda la noche (12 o más horas)

Funcionalización con Aminosilano y Epoxisilano:

Se usan los portaobjetos lavados con NaOH·EtOH

1. 5 min. En una solución 1% γ -aminopropiltrimetoxisilano 1% glicidoxipropoxitrimetoxisilano en 95% acetona
2. 2x 2 min. En acetona
3. 2x 2 min. En etanol
4. Centrifugado 3 min. a 700rpm
5. Hornear a 85°C durante toda la noche (12 o más horas)

Funcionalización con Epoxisilano:

1. Se usan los portaobjetos lavados con NaOH·EtOH

2. 5 min. En una solución 1% glicidoxipropoxitrimetoxisilano en 95% acetona
3. 2x 2 min. En acetona
4. 2x 2 min. En etanol
5. Centrifugado 3 min. a 700rpm
6. Hornear a 85°C durante toda la noche (12 o más horas)

Funcionalización con Isotiocianato:

1. Se usan portaobjetos modificados previamente con aminosilano
2. 120 min. en una solución 0.02% fenildiisotiocianato p/v en 10% piridina 90% dimetilformamida
3. 2x 2 min. En metanol
4. 2x 2 min. En etanol
5. Centrifugado 3 min. a 700rpm

Tabla 9. Protocolos de funcionalización de superficies

El proceso de fabricación de las micromatrices para el estudio de la respuesta inmune humoral (Garcia-Albert *et al.*, 2003) empleó los péptidos sintéticos de interés acoplados con BSA en una solución de impresión consistente en PBS-glicerol 10% para su posterior transferencia a la superficie de portaobjetos de vidrio funcionalizados con Poly-L-Lisina (Schena *et al.*, 1995).

3.5 Reacciones de hibridación DNA:DNA sobre micromatrices

Las reacciones de hibridación entre las muestras marcadas fluorescentemente y las sondas inmovilizadas sobre la superficie de las micromatrices se realizan siguiendo el protocolo descrito en la tabla 10.

Protocolo para la realización de ensayos de hibridación

Preparación de las micromatrices

1. Se marcan con un lápiz de punta de diamante para delimitar las regiones de reacción
2. Circunstancialmente, las micromatrices pueden ser desnaturalizadas mediante 2 lavados de 3 minutos en H₂O a 95° C

Preparación de la mezcla de reacción (volumen final 14µl)

1. Se añaden 5 µl de producto marcado de PCR al buffer de hibridación de manera que la solución final sea 5xSSC y 10% formamida.
2. Desnaturalización del producto de PCR marcado (95°C 5 min. y 4°C 5 min.)
3. Se añade SDS para una concentración final de 0.1%

Reacción

1. Se aplica la mezcla de reacción sobre la superficie de la micromatriz y se cubre la misma con un trozo de cubreobjetos de vidrio de 22x22 mm
2. Se incuba la reacción durante 2 horas en cámara húmeda

Lavados

1. Lavado de 5 min. en 1xSSC-0.1%SDS.
2. Lavado de 5 min. en 0.1xSSC-0.1%SDS.
3. Aclarado en 0.1xSSC
4. Aclarado en EtOH 95%.
5. Secado mediante centrifugación.

Las micromatrices una vez lavadas están listas para su escaneado.

Tabla 10. Protocolo de las reacciones de hibridación

3.6 Reacciones de extensión de cebador sobre micromatrices

En el caso de las reacciones de extensión de cebador, se emplea como muestra un producto de PCR sin marcar fluorescentemente. Las condiciones de las reacciones de

extensión de cebador están ligeramente modificadas respecto a las descritas en (Kurg et al., 2000). Las condiciones empleadas están descritas en la tabla 11.

Protocolo para la realización de ensayos de extensión de cebador

Preparación de las micromatrices

1. Se marcan con un lápiz de punta de diamante para delimitar las regiones de reacción
2. Se desnaturalizan las matrices mediante 2 lavados de 3 minutos en H₂O a 95° C
3. Se sitúan las matrices sobre la placa calefactada a 60° C

Preparación de la mezcla de reacción

1. Se parte de tres soluciones iniciales:
Solución A contiene 5 µl de 10x Thermosequenase reaction Buffer + 5 µl de H₂O.
Solución B contiene Thermosequenase 5U/ µl.
Solución C contiene una mezcla con los 4 dNTPs a 50 µM (incluyendo el dNTP marcado fluorescentemente)
2. A 10 µl de *Solución A* se añaden 5 µl de producto de PCR purificado
3. Se desnaturaliza la muestra durante 10 min. a 95° C y 4 min. en hielo
4. Durante el proceso de desnaturalización se mezclan 4µl *Solución C* con 1µl de *Solución B*.
5. Tras la desnaturalización se añade la mezcla B+C al productor desnaturalizado

Reacción

1. Se aplica la mezcla de reacción sobre la superficie de la micromatriz y se cubre la misma con un trozo de parafilm de aproximadamente 22x22 mm
2. Se incuba la reacción durante 25 minutos a 60° C

Lavados

1. Se realizan 2 lavados de 3 minutos con H₂O a 95° C.

En caso de quedar algún resto de agua sobre la superficie de las matrices es retirado con una corriente de aire a presión o papel con cuidado de no tocar la superficie donde se ha dado la reacción.

Las micromatrices una vez lavadas están listas para su escaneado.

Tabla 11. Protocolo y condiciones para los ensayos de extensión de cebador

3.7 Reacciones antígeno-anticuerpo sobre micromatrices

Las reacciones realizadas sobre la superficie de las micromatrices para el análisis de la respuesta inmune humoral fueron reacciones de tipo antígeno anticuerpo entre los péptidos inmovilizados sobre la superficie de las micromatrices y los anticuerpos presentes en las muestras analizadas. El protocolo seguido para estas reacciones está basado en protocolos de ensayos ELISA descritos en (Garcia-Albert et al., 2001). En este caso se emplean sueros diluidos 1/100 que se incuban durante 1 hora a 40°C en una cámara húmeda. Una vez realizada esta incubación se realizan 3 lavados de 10 minutos empleando PBS-Tween 20 0.5%, seguidos de un aclarado en PBS y posterior centrifugación para secar las micromatrices. Una vez secas se procede a la incubación con el conjugado marcado fluorescentemente con una concentración de 2.5 µg/µl durante 30 minutos a 40°C en cámara húmeda. Se realiza un nuevo ciclo de lavados con PBS-Tween 20 al 0.5%, seguidos del aclarado y la centrifugación para el secado, quedando de este modo listos para su escaneado.

3.8 Escaneado, análisis de las imágenes y análisis estadístico de los datos

Las micromatrices, una vez realizadas las reacciones de afinidad sobre su superficie, ya están listas para su escaneado. El proceso de escaneado se realizó empleando un escáner láser confocal Affymetrix-418 (Affymetrix, Santa Clara, CA, EE.UU.) o ScanarrayHT (Perkin-Elmer, Woburn, MA, EE.UU.). Las imágenes generadas durante este proceso fueron posteriormente analizadas en los casos que así lo requerían empleando el software de análisis de imágenes ImaGene 4.0 o ImaGene 6.0 (Biodiscovery, Sunnydale, CA, EE.UU.).

Los datos de estos análisis fueron guardados en AMANDA (Lopez-Campos *et al.*, 2006) una base de datos que cumple con los estándares de MIAME (Brazma *et al.*, 2001).

Los datos generados en todo este trabajo han sido analizados estadísticamente empleando diferentes programas, MS Excel (Microsoft, Redmond, WA, EE.UU.), G-Stat (GSK, Madrid) y BRB-Array Tools (Simon and BRB-ArrayTools Development Team, 2001)

3.9 Secuenciación y análisis filogenético

En el análisis de las muestras procedentes de virus respiratorios fue necesario proceder a reacciones de secuenciación para la validación de los resultados obtenidos mediante las micromatrices. Estas reacciones de secuenciación se llevaron a cabo mediante la purificación de los productos de la reacción de NESTED PCR identificados en un gel de agarosa utilizando Quiaquick PCR purification Kit (Qiagen, Hilden, Alemania) y siguiendo la instrucciones del proveedor. Una vez purificados, estos productos fueron secuenciados empleando el BigDye Terminator Cycle Sequencing v2.0 Ready Reaction Kit (Applied Biosystems, Foster City, CA, EE.UU.). Para esa reacción de PCR se utilizaron los mismos cebadores empleados en la reacción de NESTED PCR a una concentración de 3.2pmoles por 100ng de DNA molde. Una vez generadas las secuencias, la asignación de las mismas a los diferentes serotipos de los adenovirus se realizó mediante un análisis filogenético empleando el programa Molecular Evolutionary Genetic Analysis (MEGA) versión 3.1. (Kumar *et al.*, 2004) usando 10.000 réplicas de bootstrap.

4. Resultados

4.1 Sistemas de detección de patógenos humanos basados en micromatrices de material biológico:

Se han desarrollado dos aproximaciones destinadas a la validación de las micromatrices como sistemas de detección de patógenos humanos. En la primera de ellas se han desarrollado metodologías dirigidas a la detección de patógenos de interés en bioterrorismo, lo que ha incluido el diseño, puesta a punto y validación con muestras de laboratorio de los métodos de detección. En la segunda de las aproximaciones se ha optado por el análisis y desarrollo de sistemas para la detección de virus respiratorios empleando para ello muestras clínicas disponibles en el Servicio de Virología del ISCIII. En ambas aproximaciones fue necesario el desarrollo de técnicas de biología molecular para la amplificación, marcaje y detección de las muestras analizadas, empleándose para la amplificación de las muestras en algunos casos reacciones de PCR anteriormente descritas en la literatura y en otros siendo necesario el diseño y puesta a punto de las mismas.

4.1.1 Fabricación de las micromatrices.

Las micromatrices empleadas en todos estos trabajos han sido desarrolladas en el propio laboratorio en todas sus etapas, desde el diseño de las sondas inmovilizadas hasta su fabricación en la Unidad de Bioinformática empleando para ello superficies modificadas y preparadas en el propio laboratorio. El diseño espacial de las micromatrices empleadas ha ido variando con el tiempo a medida que se diseñaban nuevas sondas o ajustándose a los requerimientos solicitados por grupos con los que se ha colaborado. En esta sección se van a presentar los resultados obtenidos durante estas etapas.

4.1.1.1 Funcionalización y análisis de superficies

El soporte clásico para la fabricación de las micromatrices son los portaobjetos de vidrio. En condiciones normales, el vidrio empleado presenta poca afinidad por las moléculas biológicas de interés que se desea inmovilizar por lo tanto es necesario modificar las propiedades de las superficies de los portaobjetos para permitir que estos sean capaces de retener en la superficie las sondas que se desean inmovilizar. A este proceso de modificación de superficies se le conoce como “funcionalización”. Como parte del proceso de desarrollo de la metodología para el desarrollo de las micromatrices se analizó el comportamiento de varias superficies a la hora de inmovilizar el material biológico para la realización de los ensayos de hibridación. Las superficies analizadas fueron, poly-L-lisina (superficie con grupos NH₂ e interacciones electrostáticas entre sondas y superficie), aminosilano (superficie con grupos NH₂ e interacciones electrostáticas entre sondas y superficie), epoxisilano (superficie con grupos epóxido y enlaces covalentes estables entre sondas y superficie), epoxiaminosilano (superficie preparada con 50% grupos amino y 50% de grupos epóxido) e isotiocianato (superficie con grupos -SCN y enlaces covalentes estables entre sondas y superficie). El análisis comparativo entre estas superficies se llevó a cabo mediante la realización de una

serie de hibridaciones empleando las mismas condiciones y analizando de una forma visual y cuantitativa los resultados de las mismas (Ver figura 2).

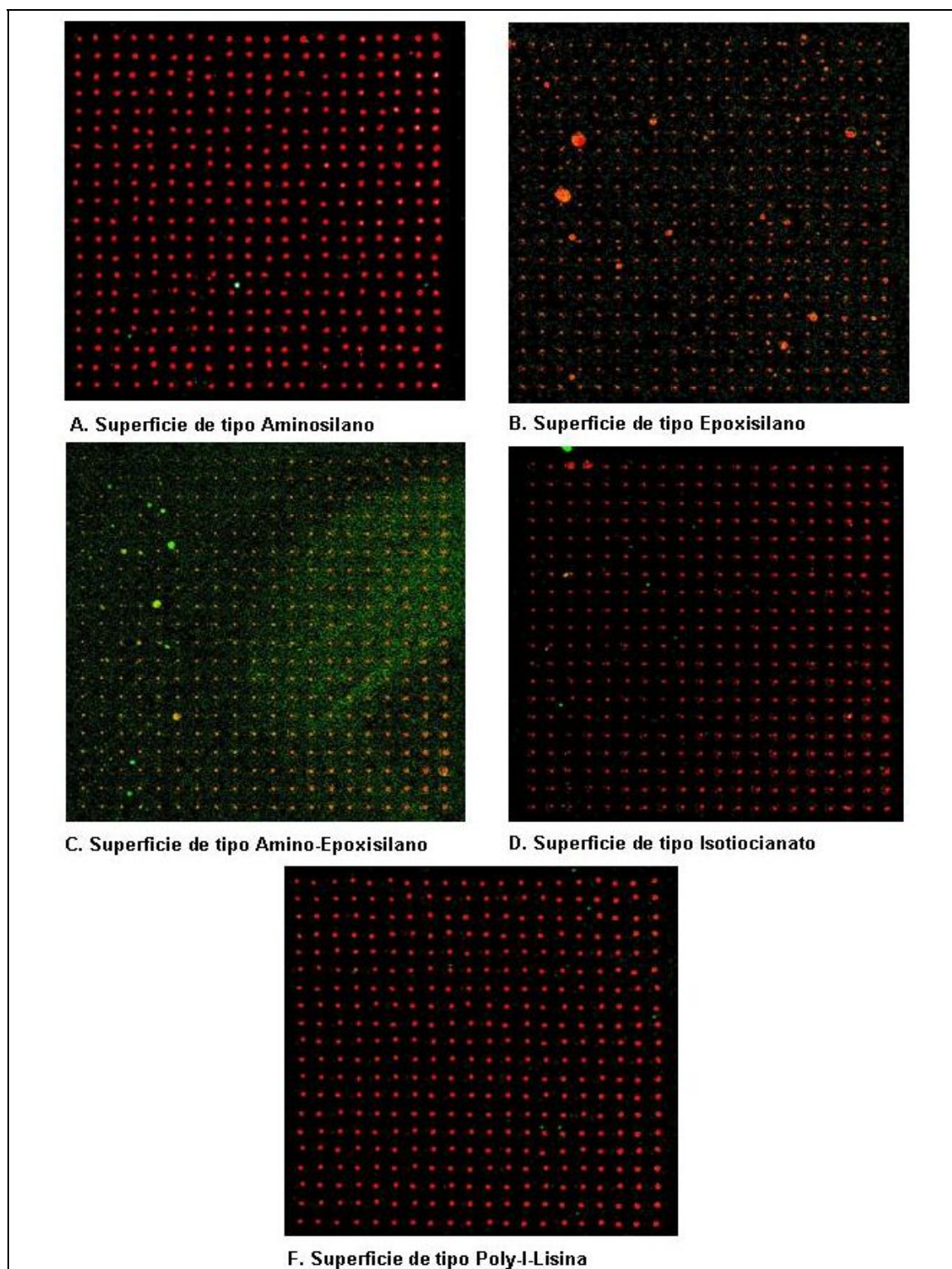


Figura 2 Imágenes de los resultados de hibridaciones de control empleando las mismas condiciones sobre las diferentes superficies analizadas. En esta figura se puede observar que visualmente hay pocas diferencias entre las diferentes superficies, siendo la diferencia más apreciable la variación en el diámetro de los puntos que se generan durante el proceso de hibridación sobre las diferentes superficies

Estos resultados mostraron que los ensayos que generaban unas mayores intensidades eran los que estaban realizados sobre una superficie de tipo epóxido pero esta superficie era también la que presentaba una mayor variabilidad en los valores generados. Las superficies que presentaban unas mejores características en cuanto a intensidades observadas, bajo ruido de fondo y reproducibilidad de los puntos generados durante el proceso de fabricación fueron las superficies de poly-L-lisina e isotiocianato. Ver figuras 3 y 4.

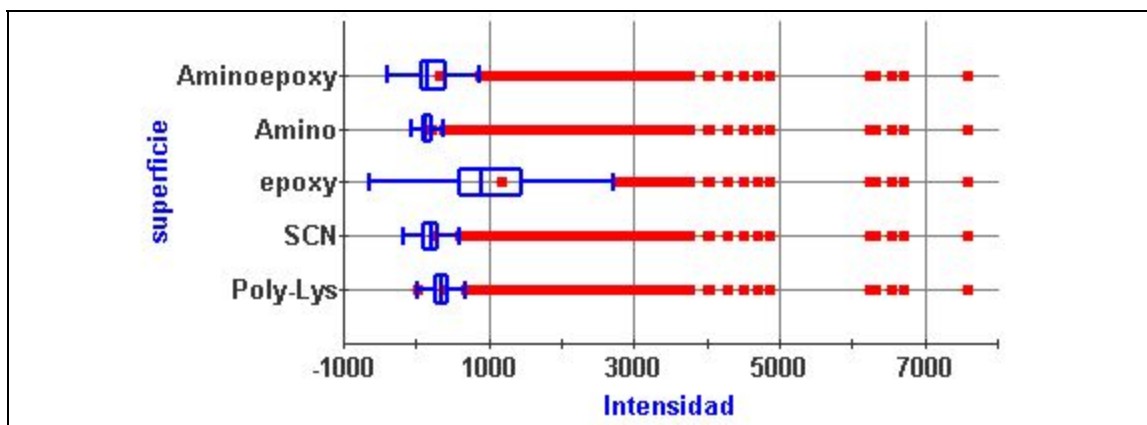


Figura 3 Gráfica con las intensidades (con el ruido de fondo corregido) generadas durante el análisis de las diferentes superficies. Se puede observar que la superficie que genera una mayor intensidad es la superficie de tipo epóxido (Epoxi) pero esta es también la que presenta una mayor dispersión. A continuación y tras la mayor intensidad de la superficie epóxido se situarían las superficies de poly-L-lisina (Poly-Lys) e isotiocianato (SCN), que además son las que presentan menores dispersiones de los datos. Por este motivo se seleccionaron estas dos superficies para la realización de los ensayos.

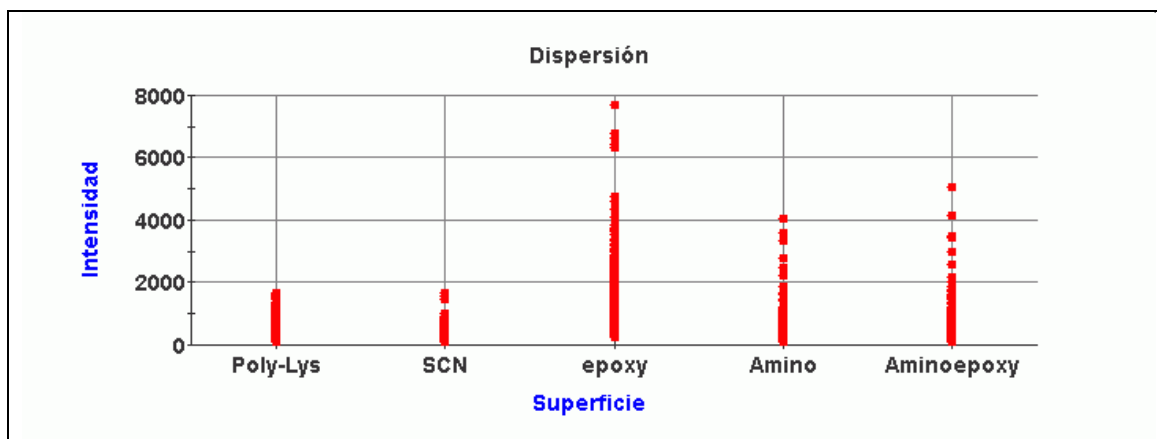


Figura 4 Gráfica con las dispersiones observadas al cuantificar las hibridaciones realizadas durante el análisis de las diferentes superficies para la inmovilización de los ácidos nucleicos. Se puede observar que las superficies que presentan un mejor comportamiento son las superficies poly-L-lisina (Poly-Lys) e isotiocianato (SCN)

Los primeros ensayos realizados se basaron en la utilización de superficies con grupos amino para la inmovilización de las sondas de oligonucleótidos mediante interacciones electrostáticas. Este tipo de aproximación es válida para la realización posterior de ensayos de hibridación en los que las condiciones de orientación de la sonda sobre la superficie no son importantes, además este tipo de inmovilización permite la utilización de sondas que no requieren de ningún tipo de modificación. Posteriormente para la realización de ensayos

enzimáticos de extensión de la sonda inmovilizada fue necesario modificar el tipo de superficie a emplear dado que en este caso se requiere una presentación de la sonda orientada en el espacio, inmovilizándose a través de su extremo 5' y manteniendo su extremo 3'-OH libre para la reacción de extensión. La superficie seleccionada para la realización de este tipo de ensayos fue de tipo isotiocianato. Este tipo de inmovilización requiere de la modificación de las sondas en su extremo 5' con la inclusión de un grupo -NH₂ para que reaccione con la superficie de la micromatriz.

4.1.1.2. Diseño de sondas para la detección de los organismos de interés

El proceso de diseño de las sondas es una etapa crucial para el desarrollo de los ensayos sobre las micromatrices ya que en él se definen los elementos responsables de la detección y que van a reaccionar con las muestras.

El proceso de diseño de sondas es un proceso bioinformático que esta constituido por varias etapas y que se define en función del tipo de ensayos que se desee realizar. Las características de las sondas son muy variables, dependiendo del entorno de aplicación en el que se desee trabajar y del tipo de ensayos que se desee realizar. Para ello se ha desarrollado un protocolo online de ayuda a la toma de decisiones para el diseño de sondas. Esta herramienta se ha diseñado para permitir tanto el diseño de sondas destinadas a la realización de ensayos de expresión génica como para ensayos en los que se deseen realizar estudios de detección de mutaciones, resecuenciación u otras aproximaciones dirigidas al análisis de secuencias de ADN (ver figura 5) y en su versión online (http://biotic.isciii.es/paginas/diseño%20de%20sondas_2.htm) tiene links a algunos de los recursos de libre acceso disponibles para realizar las etapas detalladas.

Dentro de la fase experimental de diseño de sondas, se han desarrollado sondas para dos proyectos diferentes: las que permiten la detección de virus DNA y bacterias para las micromatrices destinadas a la detección de agentes biológicos de interés en bioterrorismo y las que permiten detectar virus RNA y virus DNA en el caso de las aplicaciones para la detección de virus respiratorios en muestras clínicas. Así mismo se han desarrollado sondas para la detección por reacciones de hibridación entre las muestras y las sondas y mediante la realización de ensayos enzimáticos sobre la superficie de las micromatrices en los que las sondas actúan como sustrato.

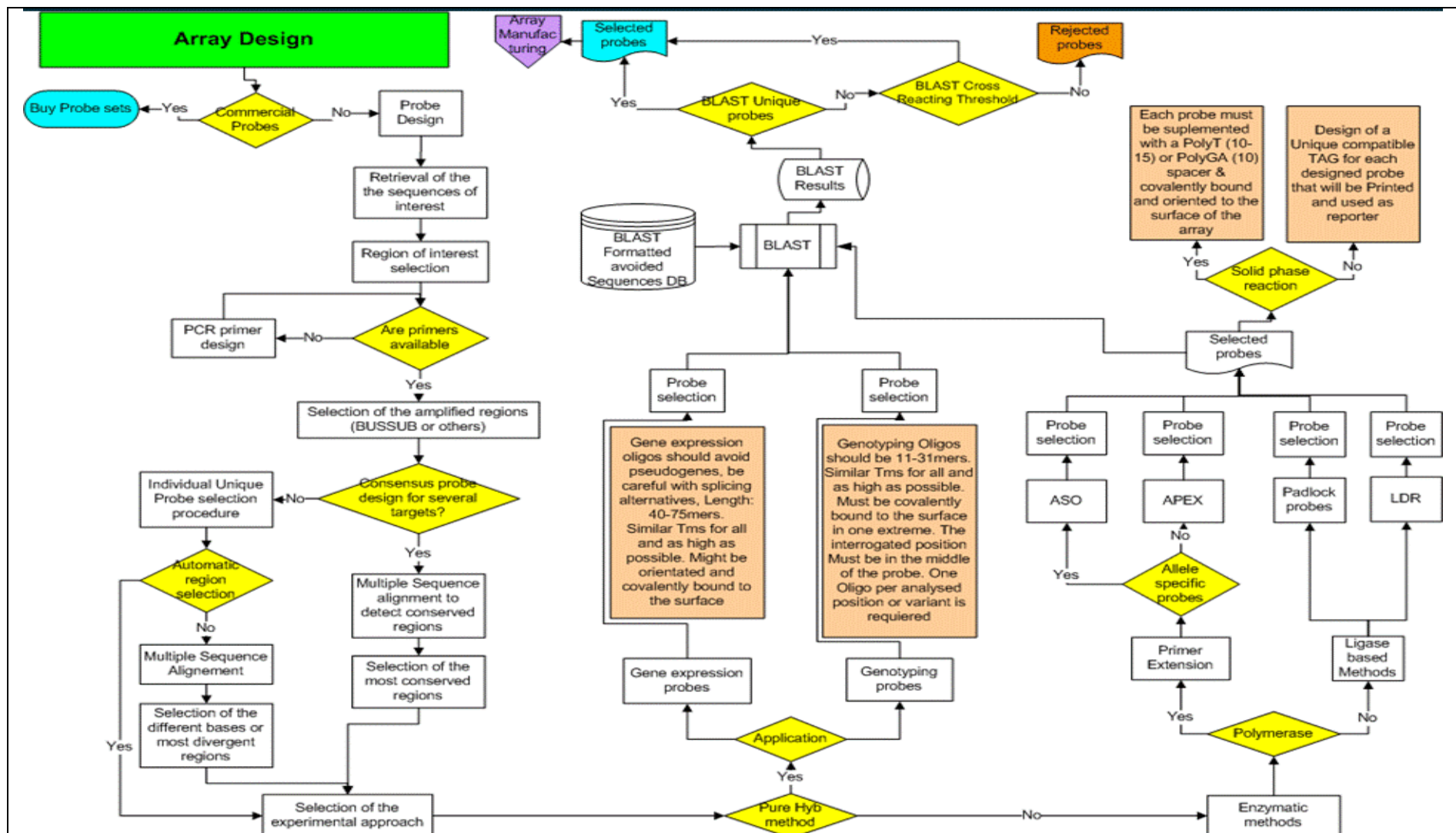


Figura 5 Esquema del sistema para la ayuda a la toma de decisiones en el proceso de diseño de sondas.

4.1.2 Sistema para el análisis de reacciones de PCR “*in Silico*” y pretratamiento de datos para el diseño de sondas de ácidos nucleicos (BUSSUB)

BUSSUB es un programa parametrizable diseñado para la búsqueda y recuperación de amplicones a partir de un conjunto de secuencias limitados por un par de secuencias flanqueantes (tales como un par de cebadores de PCR). Se han desarrollado dos versiones de BUSSUB, una diseñada como una versión para su instalación en local y que utiliza la línea de comandos y una versión online con un interfaz Web que se ejecuta en un servidor (<http://ibsen.tic.udc.es/bussubWeb2>).

La herramienta ha sido desarrollada para ser empleada fundamentalmente en un entorno de microbiología clínica en el cual los cebadores usados para las reacciones de amplificación pueden contener posiciones degeneradas. Por este motivo el programa es capaz de reconocer y aceptar posiciones degeneradas en los cebadores cuando estas estén codificadas empleando el código IUPAC. En el entorno microbiológico para el que ha sido desarrollada la herramienta, la variabilidad de las secuencias es una característica inherente a muchos de los organismos (por ejemplo virus RNA). Estos cambios son desconocidos para los usuarios en muchas ocasiones y son diferentes a los recogidos mediante la inclusión de posiciones degeneradas en los cebadores. Para hacer frente a esta variabilidad desconocida la herramienta permite al usuario definir un número determinado de apareamientos incorrectos independientes de los recogidos mediante la inclusión de degeneraciones que pueden darse a la hora de reconocer una secuencia flanqueante. Este número de variaciones es independiente para cada una de las secuencias flanqueantes.

El sistema permite la definición de una zona de especificidad de la polimerasa en el extremo 3' de la secuencia flanqueante. Este parámetro definido por el usuario recoge el número de posiciones en el extremo 3' de la secuencia en el cual no se permite la inclusión de ningún apareamiento incorrecto. Gracias a esta posibilidad el usuario puede definir cuantas bases deben ser reconocidas específicamente por la polimerasa para que esta pueda llevar a cabo la extensión. Además, cualquier otra posición fuera de este rango será considerada como una mutación que puede afectar a la eficiencia de la reacción de PCR que se desee realizar pero no evita la extensión del cebador. Finalmente los parámetros “Upstream” y “Downstream” disponibles en la entrada de datos, permiten la recuperación de regiones externas respecto a las definidas por las regiones flanqueantes.

BUSSUB genera como resultado cuatro archivos que contienen diferente información: L-file, que contiene la mayor región definida por las regiones flanqueantes con un apareamiento perfecto que se ha encontrado en las secuencias analizadas; M-File, es un archivo que contiene en formato FASTA las subsecuencias encontradas, incluyendo subsecuencias que puedan tener uno o más apareamientos incorrectos con las regiones flanqueantes que cumplen los parámetros definidos por el usuario; PM-File, es un archivo que contiene en formato FASTA las subsecuencias que se han recuperado que tienen un apareamiento perfecto con las

regiones flanqueantes; y el D-File que es un archivo con un formato tabular en el que se encuentran todas las secuencias recuperadas (con y sin apareamientos incorrectos) , incluyendo una breve descripción de la subsecuencia recuperada como la longitud y el % GC. En el caso de que las subsecuencias recuperadas contengan uno o más apareamientos incorrectos, este dato es incluido también como un descriptor en este archivo (ver figura 6.)

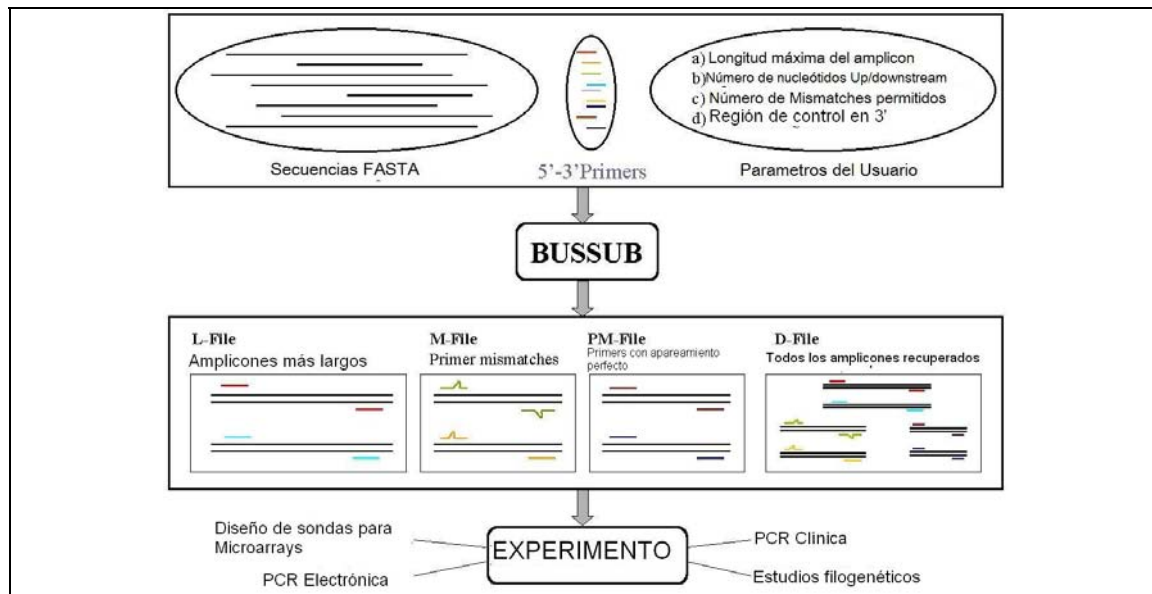


Figura 6. Esquema del funcionamiento de BUSSUB en el que se muestran los parámetros de entrada y los archivos de salida

Para validar esta herramienta se realizaron varios análisis. La primera categoría de análisis consistió en la utilización de diversos archivos multisequencia de diferentes tamaños y complejidad para analizar la capacidad computacional del sistema. Los resultados de estos análisis mostraron que el sistema es capaz de procesar grandes archivos (superiores a 200Mb) en tamaño y con secuencias de gran complejidad y tamaño como pueden ser genomas bacterianos completos o cromosomas humanos.

Una vez validado computacionalmente el sistema, la siguiente etapa consistió en su validación mediante la aplicación en problemas biológicos reales. Para ello se empleó inicialmente la reacción de PCR para la amplificación de adenovirus (Coiras *et al.*, 2003) para seleccionar las regiones amplificadas que posteriormente sirvieron para el desarrollo de las sondas usadas en la discriminación entre los diferentes serotipos. Este análisis se basó en la utilización de unos cebadores muy conservados sobre virus ADN que permitieron la detección de todos los productos de amplificación esperados y generados mediante el análisis empleando el programa BLAST, observándose además que la reacción de PCR es muy específica y genera un único producto.

Posteriormente a este análisis para la detección de amplicones en virus DNA se seleccionó una PCR clínica descrita para la detección de virus dengue (Holmes and Burch, 2000; Rico-Hesse, 2003). Este modelo fue seleccionado debido a su complejidad, dado que los virus dengue son virus RNA que presentan una gran variabilidad en sus secuencias. De este

modo, se pudo comprobar la capacidad de BUSSUB para trabajar con cebadores degenerados y las opciones de apareamientos incorrectos. Este análisis mostró que BUSSUB fue capaz de recuperar todas las secuencias esperadas debido a su capacidad para trabajar con oligonucleótidos degenerados y además fue capaz de predecir productos de amplificación inesperados que pueden ser detectados mediante geles de electroforesis y que no se pueden predecir mediante los análisis que emplean BLAST debido a la variabilidad en las secuencias virales (Domingo *et al.*, 2004).

4.1.3 Reacciones de marcaje de las muestras:

El proceso de marcaje fluorescente de las muestras destinadas a reacciones de hibridación sobre las micromatrices se incluyó en la etapa de amplificación de las mismas mediante la adición de un análogo de nucleótido marcado con un grupo fluorescente.

Durante el proceso de marcaje de las muestras se estudió el efecto que la incorporación de un análogo de nucleótido marcado tenía sobre las reacciones de PCR. Para ello se analizó el efecto que tenía la sustitución en diferentes proporciones del nucleótido sin marcar por el análogo marcado. Se pudo apreciar que se produce una caída en la eficiencia de la reacción de PCR a medida que el porcentaje de nucleótido marcado aumenta (ver figuras 7, 8 y 9). Por este motivo se seleccionó la relación 1/10 (nucleótido marcado/ nucleótido sin marcar) para la realización de todos los marcajes.

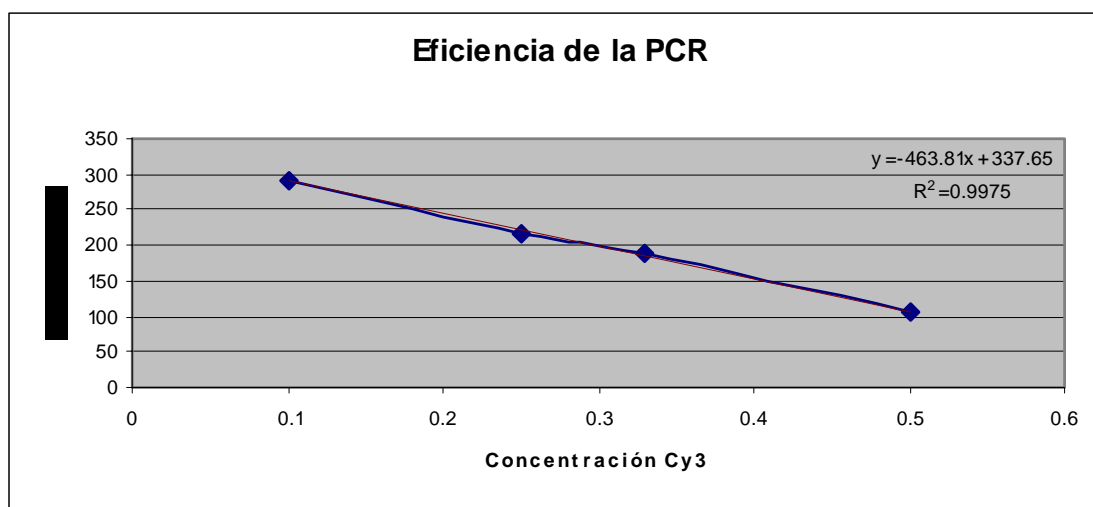


Figura 7 Recta ajustada de la caída de la eficiencia en la reacción (producto final amplificado) de PCR debida a la incorporación de diferentes proporciones de un análogo de nucleótido marcado fluorescentemente.

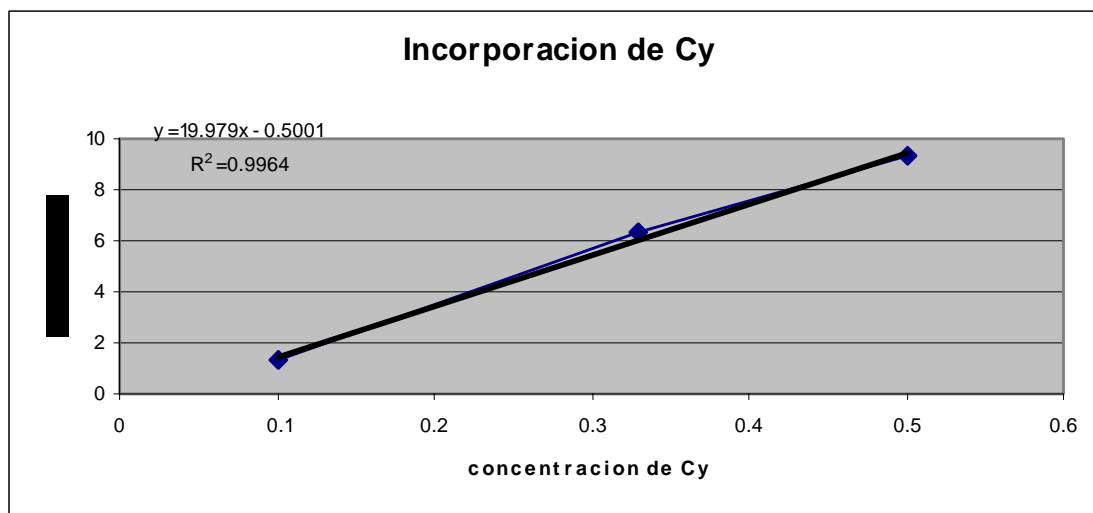


Figura 8 Recta ajustada de la incorporación de marcador fluorescente en los productos de las reacciones de PCR cuando se incluyen en la reacción diferentes proporciones de un análogo de nucleótido marcado fluorescentemente.

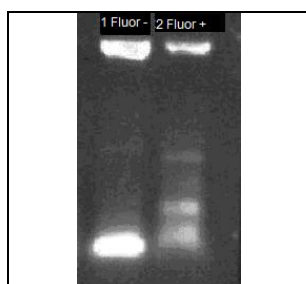


Figura 9 Diferencias en la cantidad de producto observada al correr un gel de agarosa con muestras marcadas sin usar un análogo de nucleótido y un análogo de nucleótido marcado. En el carril 1 muestra un producto marcado usando un primer que tiene incluido Cy-3 mientras que en el carril 2 la muestra se marcó con un análogo de nucleótido (dCTP-Cy-3). Se observa una menor eficiencia en la reacción 2.

En los ensayos de hibridación en los que se emplearon estos productos marcados no se detectó ninguna diferencia en la eficiencia de la hibridación en el ruido de fondo de la misma en función de que la muestra hubiese sido purificada previamente o no.

En las reacciones de extensión sobre la superficie de la micromatriz se emplean muestras amplificadas sin marcar dado que el proceso de marcaje se lleva a cabo durante la reacción de extensión de la sonda inmovilizada sobre la superficie empleando análogos de nucleótido marcados fluorescentemente.

4.1.4 Sistemas de detección de agentes susceptibles de ser empleados en bioterrorismo

4.1.4.1 Diseño y generación de un DNA quimérico de variola minor.

En colaboración con el grupo del Instituto de Salud Carlos III dirigido por el Dr. A. Tenorio se ha trabajado en el desarrollo de las sondas y sistemas de detección para orthopoxvirus. Debido a restricciones existentes para obtener muestras para la realización de los ensayos de validación, se ha desarrollado una muestra quimérica sintética diseñada para servir como control positivo.

Para el desarrollo y diseño de los oligonucleótidos de la secuencia quimérica se ha seleccionado la secuencia Y16780 perteneciente al genoma completo de variola minor y dentro de esta se ha seleccionado únicamente la región del gen crmB para la cual el grupo del Dr. A.

Tenorio tiene puesta a punto una nested-PCR que permite su detección. Este amplicón que se genera tiene un tamaño de 290pb. La metodología de diseño de las sondas quiméricas se ha desarrollado en colaboración con el laboratorio del Dr. A. Tenorio y ha sido realizada empleando herramientas bioinformáticas que han permitido seleccionar diferentes pares de oligonucleótidos que poseen regiones solapantes que permiten su elongación, para generar finalmente la secuencia de interés.

Para la generación de la sonda quimérica se ha seguido una aproximación anidada para la elongación de la secuencia usándose diferentes pares de oligonucleótidos que permiten el crecimiento de la secuencia. Así, a partir de una primera pareja de oligonucleótidos que presenta una región solapante que se elonga se genera una secuencia que sirve de molde para el siguiente par de oligonucleótidos. De esta manera y empleando un pequeño número de reacciones de PCR se puede generar la secuencia quimérica deseada. Ver figura 10.

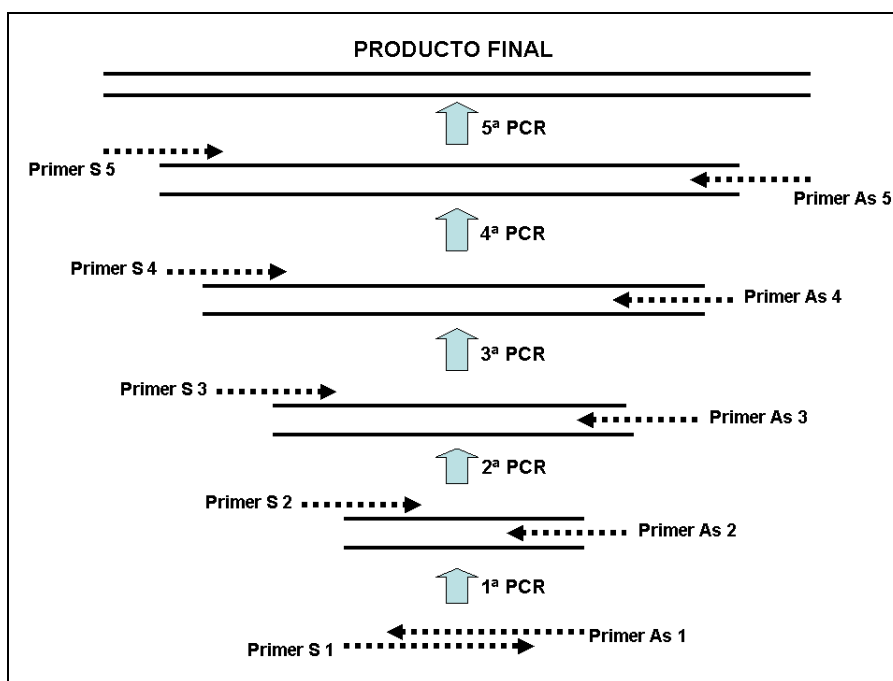


Figura 10 Esquema del proceso de generación de la quimera, en el que se ve el crecimiento mediante varias reacciones de PCR empleando oligonucleótidos con regiones que se solapan entre ellos.

Dado el tamaño del amplicón deseado (290pb), para el diseño de los oligonucleótidos se ha realizado un abordaje en el que se han tenido en cuenta las eficiencias de síntesis así como la minimización del número de reacciones y etapas necesarias para la generación de la secuencia quimérica. Para minimizar los posibles errores en la síntesis de la secuencia generada respecto a la secuencia deseada se han seleccionado oligonucleótidos que no tuviesen un tamaño demasiado largo. Los oligonucleótidos se han diseñado de manera que las regiones de los mismos que deben solaparse y van a servir como cebadores para las reacciones de extensión de los mismos tuvieran una temperatura de anillamiento similar. De esta forma se logra que todas las reacciones necesarias para la extensión y generación de la secuencia final tengan un comportamiento similar y se puedan realizar bajo las mismas

condiciones experimentales. Por estos motivos se diseñaron 5 pares de oligonucleótidos que tienen unos tamaños que oscilan entre 58pb y 40pb, que presentan unas regiones solapantes de alrededor de 20 bases. En la tabla 12 se presentan los oligonucleótidos diseñados y que sirven como primers sense y antisense en cada una de las reacciones de PCR.

Nombre	Secuencia 5'-3'	Longitud	Pos3'	Pos5'
S5	TAT TAC TAT GAA TCA TAC AGA TTG CAA TCC CGT ATT TCG TGA GGA ATA CTT CTC T	55	1	55
S4	TAT TTC GTG AGG AAT ACT TCT CTG TCC TTA ATA AGG TAG CAA CTT C	55	33	78
S3	GTC CTT AAT AAG GTA GCA ACT TCA GGA TTT TTT ACA GGA GAA AAT AGA TAT CA	52	56	108
S2	TTT ACA GGA GAA AAT AGA TAT CAA AAT ATT TCA AAG GTG TG	40	86	126
S1	TAT CAA AAT ATT TCA AAG GTG TGT ACT TTA AAT TTT GAG ATT AAA TGT AAT AAC AAA GG	58	104	162
As1	TTC GTT AGC TGT TTG AAG GAA GAA CCT TTG TTA TTA CAT TTA ATC TC	47	140	187
As2	GAC ATC ATA CCG TCA TCA TTC TTT GCT TTC GTT AGC TGT TTG AAG G	45	168	213
As3	CGC TAG AGT TAC CGT CTC CGA ATG CGA CAT CAT ACC GTC ATC AT	43	195	238
As4	AGA TGT CGA CGC TAG ATA GAC AGT CAC CCG CTA GAG TTA CCG TCT C	45	221	266
As5	CAT TGG TAT TAC TAT ATA GTA TAT AGA TGT CGA CGC TAG ATA GAC	44	246	290

Tabla 12 Oligonucleótidos sintéticos empleados para la generación de un control positivo para variola minor con sus características.

Empleando esta aproximación se logra elongar la secuencia alrededor de 60pb en cada reacción alcanzándose la secuencia deseada en tan sólo 5 reacciones de PCR. También se han realizado ensayos en los que, aprovechando que los oligonucleótidos han sido diseñados para trabajar todos en el mismo rango de temperaturas, se ha realizado una única reacción de PCR en la que también se logra generar el fragmento de amplificación deseado. Figura 11.

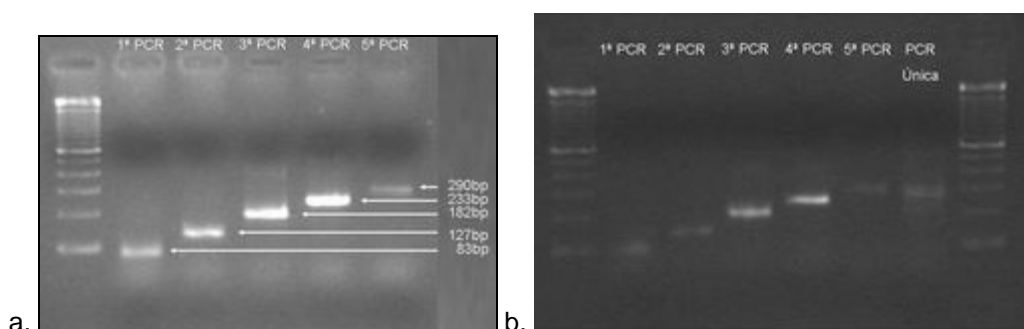


Figura 11 Gel de agarosa con los productos de las diferentes reacciones de amplificación empleadas para la generación de un DNA quimérico de variola. A. Se puede apreciar el tamaño creciente de las bandas a medida que se van empleando oligonucleótidos cada vez más externos, hasta alcanzar el tamaño deseado de 290pb que se obtiene en la 5ª y última reacción. B. Gel en el que se han corrido las mismas muestras anteriores pero en el que en la 6ª calle se ha incluido el producto generado mediante una única reacción de PCR empleando los 10 oligonucleótidos diseñados.

Este producto fue clonado y secuenciado comprobando que se trataba únicamente de la secuencia deseada (ver figura 12) para emplearse con posterioridad en los ensayos de validación de las sondas.

secuenciaciónsense	-----	0
amplicon	tattactatgaatcat-----ACAGATTGCAATCCCGTATTTCGTGAGGAATACTTCTCT	55
as1	tttttttnttntnngggatnaACAGATTGCAATCCCGTATTTCGTGAGGAATACTTCTCT	60
secuenciaciónsense	-----TTCGGGATTCTNCCGGGAGAAAATAGATATCAAAATATT	40
amplicon	GTCCTTAATAAGGTAGCAACTTCAGGATTTTTACAGGAGAAAATAGATATCAAAATATT	115
as1	GTCCTTAATAAGGTAGCAACTTCAGGATTTTTACAGGAGAAAATAGATATCAAAATATT	120
secuenciaciónsense	TCAAAGGTGTGACTTTAAATTTTGAGATTAAATGTAATAACAAAGGTTCTTCCTTCAA	100
amplicon	TCAAAGGTGTGACTTTAAATTTTGAGATTAAATGTAATAACAAAGGTTCTTCCTTCAA	175
as1	TCAAAGGTGTGACTTTAAATTTTGAGATTAAATGTAATAACAAAGGTTCTTCCTTCAA	180
secuenciaciónsense	CAGCTAACGAAAGCAAAGAATGATGACGGTATGATGTCGCATTTCGGAGACGGTAACTCTA	160
amplicon	CAGCTAACGAAAGCAAAGAATGATGACGGTATGATGTCGCATTTCGGAGACGGTAACTCTA	235
as1	CAGCTAACGAAAGCAAAGAATGATGAGGNNAAAGANNn-----	218
s1	GCGGGTGACTGTCTATCTAGCGTCGACATCTATATACTATTg-----	202
amplicon	GCGGGTGACTGTCTATCTAGCGTCGACATCTATATACTATatagtaataaccaatg	290
as1	-----	218

Figura 12 Alineamiento del amplicón esperado y las secuencias sense y antisense obtenidas tras la secuenciación de los productos amplificados.

4.1.4.2 Reacciones de amplificación para los organismos de interés:

En la mayoría de los casos las muestras se han amplificado empleando reacciones de PCR anteriormente descritas en la literatura. En el caso de las PCRs para la amplificación de las muestras para la detección de virus respiratorios, se utilizó una PCR anidada (Coiras *et al.*, 2003) basada en unos cebadores capaces de amplificar una región codificante de la proteína hexon que permite la detección de cualquier adenovirus independientemente del serotipo del que se trate.

En el caso de las reacciones de amplificación para la detección de agentes de bioterrorismo se han empleado reacciones de amplificación descritas en la literatura, transferidas por parte de grupos colaboradores o que han tenido que ser desarrolladas “ad hoc” (ver tabla 6). Las reacciones de amplificación “ad hoc” se han desarrollado para la amplificación y detección de bacterias de los géneros Burkholderia y Brucella.

4.1.4.2.1 Reacciones para la detección del género Burkholderia

El desarrollo de reacciones de amplificación para la detección de bacterias pertenecientes al género Burkholderia se llevó a cabo sobre los genes *gyrA* y *lysE*. La selección de estos genes se realizó mediante el análisis genómico de las diferentes especies pertenecientes a este género empleando los datos y herramientas disponibles a través del CMR (Comprehensive Microbial Resource, <http://cmr.jcvi.org>), (Peterson *et al.*, 2001b).

El proceso de selección de genes que pudieran ser empleados como dianas de amplificación estuvo dirigido por un análisis bioinformático de los genomas completos contenidos en la base de datos del CMR. Este análisis se dirigió hacia la detección e identificación de regiones que pudieran presentar variabilidad en las secuencias entre las diferentes especies pertenecientes al género Burkholderia. El análisis bioinformático empleando la herramienta de comparación de genomas del CMR permitió la identificación de los genes citados anteriormente. En el caso de la subunidad A de la girasa el análisis de la

comparación de genomas empleando las herramientas del CMR arroja unos valores de conservación que oscilan entre el 98% y 93% entre las especies analizadas (el genoma de *B. mallei* ATCC23344 fue empleado como referencia para la comparación con *B. pseudomallei* 1710b, *B. pseudomallei* K96243, *B. thailandensis* E264, *B. cenocepacia* AU1054 y *B. sp* 383). La herramienta de comparación de genomas de este recurso se basa en el uso de comparaciones mediante BLASTP (Altschul *et al.*, 1990) a nivel de proteína en lugar de a nivel de ácidos nucleicos, por lo que las diferencias identificadas se deben únicamente al proceso de enmascaramiento de las secuencias y no a diferencias reales entre los genes, dado que el análisis de las secuencias de los ácidos nucleicos entre estos genes muestra una similitud del 100%. Por lo tanto las reacciones que se han puesto a punto son reacciones para la detección a nivel de género. Una vez seleccionadas estas regiones se procedió al diseño de los cebadores para estas reacciones empleando el programa Oligo6 (Molecular Insights, USA). Los resultados de los procesos de búsqueda de cebadores se presentan en la tabla 13. Estos cebadores fueron validados "*in Silico*" empleando para ello los programas BLAST (Altschul *et al.*, 1990) y BUSSUB (Lopez-Campos *et al.*, 2007).

Nombre	Secuencia	Diana	Ampl.	Lon.	Tm (°C)	Sentido
gyrA_F	GTACTCGGACAGCGGCTGCGA	gyrA	314	22	70	Sense
gyrA_R	ACGCAGGCGCAGGAAATCCTC			22	68	Antisense
lysE_F	CTTTCGCCAGAGCGTGAT	lysE	171	18	56	Sense
lysE_R	GCCGAACACCGAGAACAC			19	58	Antisense

Tabla 13 Resumen con las propiedades de los cebadores diseñados para la amplificación mediante reacciones de PCR de los diferentes genes de diferentes organismos pertenecientes al género *Burkholderia*.

Los ensayos de puesta a punto "*in vitro*" de estas reacciones de PCR se realizaron analizando diferentes condiciones de temperatura de anillamiento de los cebadores y diferentes concentraciones de MgCl₂. Todas las reacciones de ajuste se llevaron a cabo empleando una muestra inicial procedente de DNA genómico de *B. mallei* equivalente a 250 copias de genoma. Una vez realizada la puesta a punto de las reacciones empleando este organismo se procedió a la validación de las condiciones de reacción empleando ADNs de otras especies del género *Burkholderia*. Durante el proceso de puesta a punto de las reacciones de PCR se analizaron unas condiciones similares a las de otras de las PCR disponibles para permitir la adaptación futura de estos ensayos a un formato multiplexado o en su defecto a que el mismo programa del termociclador fuera empleado para el mayor número de reacciones posibles.

En la puesta a punto de las reacciones para la detección del gen *lysE* se analizó un rango de temperaturas entre los 50°C y los 60°C y un rango de concentraciones de MgCl₂ variable entre 1.5mM, 2.5mM y 3.5mM. El resultado de estos ensayos, visible en la figura 13, muestra que la concentración de MgCl₂ es más importante que la temperatura a la hora de definir las mejores condiciones de amplificación, seleccionándose las condiciones definidas por una temperatura de anillamiento de 50.1 °C y una concentración de MgCl₂ de 1.5mM.

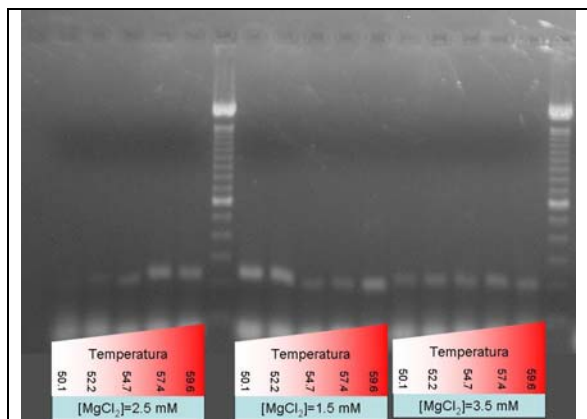


Figura 13 Imagen de un gel de agarosa en el que se han corrido los productos de amplificación de los ajustes de las condiciones de reacción para la amplificación de los productos *lysE*. En este caso se puede observar que la banda más intensa que se obtiene corresponde a la temperatura de 50.1°C y una concentración de $MgCl_2$ de 1.5mM.

Para la amplificación del gen de la subunidad A de la girasa (*gyrA*) se repitió la estrategia para la puesta a punto de las condiciones de reacción. Al igual que sucedió en el caso de la reacción para la amplificación de un fragmento del gen *lysE* se seleccionó la temperatura de 50°C dado que ofrecía una buena intensidad en las bandas analizadas, ver figura 14.

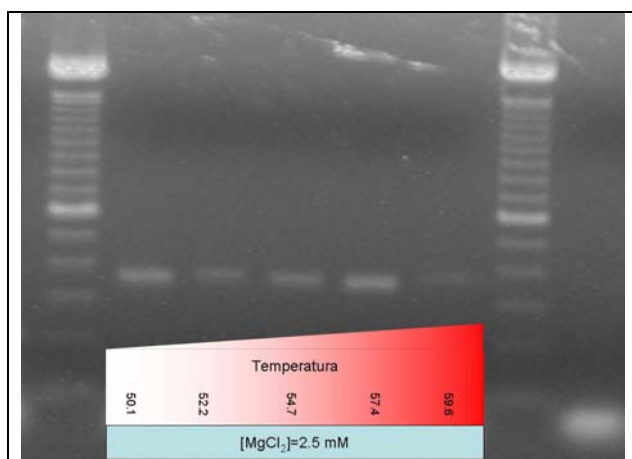


Figura 14 Imagen de un gel de agarosa en la que se puede apreciar la amplificación de *B. mallei* a diferentes temperaturas y usando una única concentración de $MgCl_2$. Se puede apreciar que la mejor señal se obtiene de la reacción realizada a 50.1 °C.

4.1.4.2.2 Reacciones para la detección del género *Brucella*.

El objetivo de estas reacciones es proporcionar material amplificado de regiones que contengan zonas altamente conservadas entre las diferentes especies del género *Brucella* que permitan, mediante ensayos sobre micromatrices, la posterior detección a nivel de género y que contengan zonas de variabilidad que permitan posteriormente discriminar entre las diferentes especies pertenecientes a este género. Para realizar este tipo de abordaje fue necesario diseñar 3 reacciones de PCR para la amplificación de los genes *bcbp31*, *omp2a/b* y *omp31*. El primero de estos genes es considerado como un gen altamente conservado en el género *Brucella* y por tanto susceptible de ser empleado para la detección de cualquier miembro de este género. El gen *omp2a/b* codifica una proteína transportadora de membrana que presenta varias regiones de variabilidad entre las diferentes especies pertenecientes al género *Brucella*. El gen *omp31*, al igual que el gen *omp2a/b* codifica una proteína transportadora de membrana y se encuentra presente en todas las especies pertenecientes al género *Brucella* con excepción de la especie *B. melitensis abortus*.

El diseño de las reacciones de PCR estuvo dirigido por la localización previa de las regiones de variabilidad, para permitir la posterior discriminación, que se deseaban amplificar.

Una vez seleccionadas estas regiones se procedió el diseño de los cebadores y puesta a punto de las condiciones experimentales necesarias para la amplificación. Al igual que se indicó anteriormente, al describir las reacciones de puesta a punto de las reacciones para la amplificación del género *Burkholderia*, se trató de ajustar el diseño de los cebadores y de las condiciones a las reacciones de PCR ya descritas para otros organismos.

El diseño de las reacciones para la amplificación de los tres genes fue muy similar, seleccionándose regiones que permitiesen el diseño de sondas discriminantes en su interior y que además presentasen la menor longitud posible. En el caso de la reacción para la amplificación se seleccionó además una estrategia basada en el desarrollo de una reacción de PCR en heminested dado que, de esta manera se lograba generar un producto de amplificación que era un mejor sustrato para las reacciones de discriminación posteriores. Para ello se desarrollaron dos cebadores sobre la hebra sentido y un cebador común para la hebra antisentido. En la tabla 14 se recogen los resultados del diseño de estos cebadores.

Cebador	Secuencia	Longitud	Tm	Horquilla ΔG	Dímero ΔG	T ^a Opt.
BCSP_S	AATAATGGAATGCGGTGGTTG	21	53,1	0	0	
BCSP_As	GTAGGCTTCAAGAACGATACG	21	52,6	-0,2	-0,2	
Producto		593	81,8			58,1
omp2_S1	CAGTTCGGTGAAGGTCTTGAG	21	64,4	1,7		61,8
omp2_As	CCGTTAAGGGCTGTTTCGAC	20	67,1	1,2		
Producto		685	92,6			
omp2_S2	TCGCCCATTATTCGTTTCAA	21	68,2	0,7		68
omp2_As	CCGTTAAGGGCTGTTTCGAC	20	67,1	1,2		
Producto		233	87,3			
omp31S	GCGGCAAGTTCAAGCATCC	19	55,6	2	-3,9	0
omp31As	GAGGTCGGTGTAGAGGTATTCC	22	55,5	2	-3,3	0
Producto		473	82,1			59,2

Tabla 14. En esta tabla se recogen las características de los cebadores seleccionados y utilizados en las reacciones de PCR para la amplificación de las muestras. Resaltar el aspecto de hemi-nested abordado para la amplificación del gen omp2

Se repitió el proceso de ajuste de condiciones como se explicó en el caso de las reacciones para la detección de *Burkholderia*. Una vez finalizadas las reacciones se realizaron los ensayos de validación empleando diferentes especies pertenecientes al género *Brucella*, obteniéndose los resultados esperados (ver figura 15).



Figura 15 Imagen de un gel de agarosa en el que se observan los resultados de la amplificación de los diferentes productos de amplificación de los genes BSCP31, omp31 y omp2 por este orden en bloques de 4. El orden de las calles en las que se han corrido las muestras con diferentes especies de *Brucella* han sido *B. melitensis abortus*, *B. melitensis melitensis*, *B. suis* y *B. canis*. Se puede apreciar que la bandas identificadas están localizadas a la altura esperada y que se produce la amplificación en las cuatro muestras con la excepción de la muestra de *B. melitensis abortus* que no tiene producto de amplificación del gen omp31 (calle 7).

Aprovechando que las reacciones se diseñaron para trabajar bajo las mismas condiciones se trató de desarrollar una reacción múltiplex con todas las reacciones para la detección de *Brucella*, pero no se logró obtener la amplificación del gene omp2a/b, de manera que finalmente se mantuvieron estas reacciones independientes y se desarrolló una reacción conjunta para la amplificación de bscp31 y omp31 (ver figura 16)

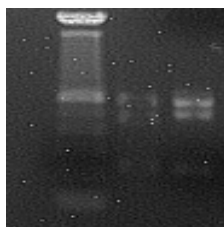


Figura 16. Imagen de un gel de agarosa en el que se ha corrido el producto de amplificación resultante del ensayo de multiplexación de las reacciones de PCR para la detección de *brucella*. Se pueden apreciar las bandas correspondientes a los productos de los genes bscp31 y omp31 únicamente. En este ensayo se han amplificado muestras de *B. melitensis melitensis* y *B. suis*.

4.1.4.3 Diseño de sondas para la detección de agentes susceptibles de ser empleados en bioterrorismo:

En el desarrollo de sistemas basados en micromatrices para la detección de agentes susceptibles de ser empleados en bioterrorismo fue necesario disponer de sondas que permitiesen la detección genérica de los organismos de interés y en muchos casos la discriminación a nivel de especie o subespecie de estos mismos organismos. Se han diseñado sondas para permitir la detección de los diferentes organismos empleando dos metodologías diferentes, una basada en ensayos de hibridación y otra basada en la extensión de sondas inmovilizadas sobre la superficie de la micromatriz.

El proceso de diseño de sondas para los ensayos de hibridación ha requerido que se aplicasen los mismos criterios y parámetros durante el diseño y la selección para permitir que todas las sondas fuesen efectivas bajo las mismas condiciones de hibridación con similar

eficiencia. Además, las sondas han sido diseñadas de manera que estuviesen presentes dentro de los fragmentos definidos por los procesos de amplificación.

Las sondas diseñadas para la discriminación de variaciones de un único nucleótido mediante ensayos de hibridación han sido diseñadas de manera que la posición analizada estuviese lo más centrada posible dentro de la sonda para maximizar los efectos estéricos que permiten la discriminación entre un apareamiento perfecto y uno imperfecto entre la muestra y la sonda.

En el caso de las sondas diseñadas para la realización de ensayos de extensión de primer se emplearon criterios propios para la realización de estos ensayos respecto a la longitud y características físico químicas de las mismas. En el caso de las sondas diseñadas para la detección de variaciones de un único nucleótido se utilizó una estrategia basada en el diseño de sondas alelo-específicas por mostrarse como una solución más económica frente a la utilización y detección de terminadores de cadena.

En la tabla 15 se recogen las sondas diseñadas para su utilización en micromatrices para la detección de agentes susceptibles de ser empleados en bioterrorismo y sus principales características.

Nombre de la sonda	Secuencia de referencia	Secuencia de la sonda	T _m (°C)	Longitud	Diana	Organismo
OvC152	M35027.1	AACAGTTAACGA AAGCAAAGAATG A	69	25	<i>Crm</i>	Vaccinia
OvM150	U87841.1	GTAACAACAAAG ATTCATCTTCCAA	67	25	<i>Crm</i>	Monkeypox
OvV119	Y16780.1	AAGGTGTGTACT TTAAATTTTGAGA	66	25	<i>Crm</i>	Variola
tul4	M32059.1	TAGGGTTAGGTG GCTCTGATGATG C	75	25	<i>tul4</i>	F. tularensis
YpestisPlasm-1251U35	M27820.1	TAGCGGGGAGG ATGTCTGTGGAA AGGAGGTTGGTG	81	35	<i>pla</i>	Y. pestis
Ypestisplasm-1351U35	M27820.1	TTGTTTTGCCTTG ACATTCTCCACG TCTTTCAGGG	77	35	<i>pla</i>	Y. pestis
YpestisPlasm-1383U35	M27820.1	GGGCATGGAGAA GGTCAAATTAGA CATGGAACGCT	78	35	<i>pla</i>	Y. pestis

Tabla 15 Conjunto con todas las sondas diseñadas para la detección de agentes susceptibles de ser empleados en bioterrorismo. Estas sondas fueron probadas en los ensayos de validación. Para la fabricación de las micromatrices se seleccionaron diferentes subconjuntos de estas sondas para su inmovilización.

Nombre de la sonda	Secuencia de referencia	Secuencia de la sonda	Tm (°C)	Longitud	Diana	Organismo
APEXasYpestisPlasm-1	M27820.1	TGAAAGACGTGGAGAATGTCAAGGC	75	25	<i>pla</i>	Y. pestis
APEXsYpestisplasm-13	M27820.1	ATCACCTTTTTCACCATAATGACGG	71	25	<i>pla</i>	Y. pestis
Ba813	U46157.1	TCCAAGTCAGAAAGAAATCCCATCA	71	25	<i>ba813</i>	B. anthracis
CapBCA	M24150.1	CTGCTTCTGTACGTTGTACCCATGTCGCAGCTATT	78	35	<i>CapBCA</i>	B. anthracis
CYA	M23179.1	TTTGCCAATTTGATAAAGTTCCCTTAGTTGAATCC	72	35	<i>cya</i>	B. anthracis
lef	M29081.1	TTGCATATTATATCGAGCCACAGCATCGTGATGTT	75	35	<i>lef</i>	B. anthracis
BantLef3/4	M29081.1	TGCAGATTCCTATTGAGCCAAAGAAAGATGACATA	73	35	<i>lef3/4</i>	B. anthracis
PAG	AE017336.2	TCCGCTGATAATCATGTAACAATGTGGGTAGATGA	74	35	<i>pag</i>	B. anthracis
BCSP 31_S_Probe001	AE009520.1	AGCCGGGTTCTGGCACCAT	78	19	<i>bcsp31</i>	Brucella melitensis 16M
BCSP 31_S_Probe002	AE009520.1	CGGGTTCTGGCATCATCGTCG	79	21	<i>bcsp31</i>	Brucella melitensis 16M
BCSP 31_S_Probe003	AE009520.1	CGTTTCGCTGGATGAGCCGGG	81	21	<i>bcsp31</i>	Brucella melitensis 16M
BCSP 31_S_Probe004	AE009520.1	TGGCTGTTGCAGCGGTGGCGG	85	21	<i>bcsp31</i>	Brucella melitensis 16M
BCSP 31_S_Probe005	AE009520.1	GCGATTGCGTTGGGAGCGAGCTTT	80	24	<i>bcsp31</i>	Brucella melitensis 16M
BCSP 31_S_Probe006	M20404.1	ATGATGGCAAGGGCAAGGTGGAAGATTT	76	28	<i>bcsp31</i>	Brucella Abortus
brucellaamp1-34U25	M20404.1	GGGCGCTCTGGAGTCCGGCTTTACG	83	25	<i>bcsp31</i>	Brucella Abortus
brucellaamp1b-137U25	M20404.1	CTGGCGACGCTTTACCCGGAAACGA	80	25	<i>bcsp31</i>	Brucella Abortus
brucellaamp2-248U25	U26440.1	CGACCGCAATGGCTGGGACAAGGGC	84	25	<i>omp2</i>	Brucella melitensis 16M
brucellaamp2-254U25	U26440.1	CAATGGCTGGGACAAGGGCGCTCGT	82	25	<i>omp2</i>	Brucella melitensis 16M
brucellaamp2-44U20	U26440.1	CTCCGCCGACGCTCTGGTTG	82	20	<i>omp2</i>	Brucella melitensis 16M
Burk1	CP000011.2	GATGAAGCTGCTCGGCACCGTGTCG	82	25	<i>Scu</i>	Burkholderia mallei
Burk2	CP000011.2	GTGCTCGCGTGCGTGGCATGGAAGA	82	25	<i>Scu</i>	Burkholderia mallei
Burk3	CP000011.2	GTGTCGGTCTGCTCGCGGCTGCTCG	85	25	<i>Scu</i>	Burkholderia mallei
BurkGyr1	CP000010.1	CGTGACCACCATGTCTGCGGCGTGAT	82	27	<i>gyrA</i>	Burkholderia mallei
BurkLys1	CP000010.1	ATCCGAAGGTGACGCTGTTCTTCGT	76	25	<i>lysE</i>	B. mallei
CoxN	M21291.1	CAGAGAGTGCAAGCGTTAATCGGAA	74	25	<i>16S RNA</i>	C. burnetii
Ftul1	Z21931.1	GGCTGCGAAGGTGCGAGCTGGAGCG	85	25	<i>16S RNA</i>	F. tularensis tularensis
Ftul16S-APEX-335U31	Z21931.1	TGGCAGCAGTGGGGAATATTGGACAATGGGGG	81	32	<i>16S RNA</i>	F. tularensis
Ftul16S-SNP-1126U25	Z21931.1	GGGTACTCTATTAAGACTGCCGCTG	75	26	<i>16S RNA</i>	F. tularensis tularensis

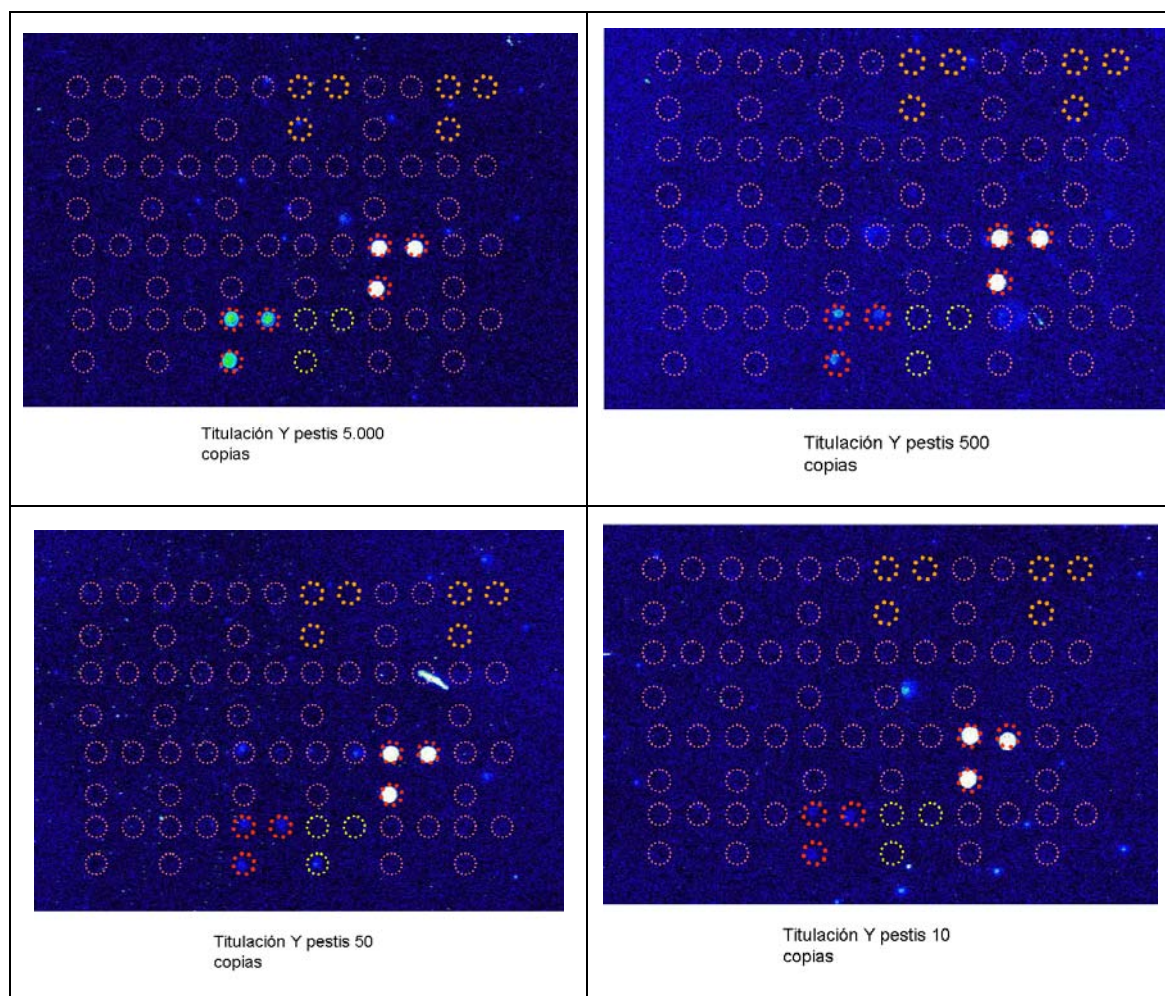
Nombre de la sonda	Secuencia de referencia	Secuencia de la sonda	Tm (°C)	Longitud	Diana	Organismo
Ftul16S-SNP-1257U27	Z21931.1	GGAGCGAAACTCAAAAAGGTACTCTTA	71	28	16S RNA	F. tularensis tularensis
Ftul2	Z21931.1	AACAGGTCTTAGGATGCTGACGAGTGGCGGA	80	31	16S RNA	F. tularensis tularensis
Ftul3	Z21931.1	GGAGGACGTTACCCAAAGAATAAGCACCGGC	80	31	16S RNA	F. tularensis tularensis
Ftul4	Z21931.1	TTTCTAGAGATAGATTGGTGCTTCGGAACGC	75	31	16S RNA	F. tularensis tularensis
Ftul5	Z21931.1	AAGGTGCGAGCTGGAGCGAAACTCAAAAAGG	79	31	16S RNA	F. tularensis tularensis
Ftul6	Z21932.1	CATTAAGTTGGGTACTCTATTGAGACTGCCGCTGACAAGG	78	40	16S RNA	F. tularensis novicida
Ftul7	Z21933.1	CGTGCCGCAGATGGATGAGCC	81	21	16S RNA	F. philomiragia
Ftul8	Z21933.1	CAGATGGATGAGCCTGCGTTGGATTAGCTAG	78	31	16S RNA	F. philomiragia
Ntul4249U30	M32059.1	ATGCTGAGAAGAACGATAAACTTGGGCAA	74	30	tul4	F. tularensis
omp2ab_S_Probe001	U26440.1	GTAAACGTCATCGCCGCCC	78	19	omp2	Brucella melitensis 16M
omp2ab_S_Probe002	U26440.1	AACGGACGTAACCATGGACG	75	20	omp2	Brucella melitensis 16M
omp2ab_S_Probe003	U26440.1	GTAACCATGGACGCGCAGG	78	19	omp2	Brucella melitensis 16M
omp2ab_S_Probe004	U26440.1	AAATGGGTCTGGGCATTCTGA	73	21	omp2	Brucella melitensis 16M
omp2ab_S_Probe005	U26440.1	CACAACCTTTACGGTGCTGCTA	74	22	omp2	Brucella melitensis 16M
omp2ab_S_Probe006	U26438.1	CACAACCTTTATGGTGCTGCTA	72	22	omp2	Brucella abortus
omp2ab_S_Probe007	U26440.1	GAAGGGCAGTTATTGCAGAAATG	72	23	omp2	Brucella melitensis 16M
omp2ab_S_Probe008	U26438.1	GAAGGGCAGTTGTTGCAGAAATG	75	23	omp2	Brucella abortus
omp2ab_S_Probe009	U26440.1	GCAGTTATTGCAGAAATGACACG	72	23	omp2	Brucella melitensis 16M
omp2ab_S_Probe010	U26443.1	GCAGTTATTGCTGAAATGACACG	72	23	omp2	Brucella suis 1330
omp2ab_S_Probe011	U26440.1	AGGATCGCCTGCCGAATAAAG	74	21	omp2	Brucella melitensis 16M
omp2ab_S_Probe012	U26443.1	AGGATCGCCTACCGAATAAAG	71	21	omp2	Brucella suis 1330
omp31_S_Probe001	CP000709.1	TTGACAAGGAAGACGGCGAACAT	74	23	omp31	Brucella ovis
omp31_S_Probe002	AE008918.1	CGGTGCCAGCGGTCTCGAA	80	19	omp31	Brucella melitensis 16M
omp31_S_Probe003	AE008918.1	GGCTTCGTCGGTGGTGTTCAGG	80	22	omp31	Brucella melitensis 16M
omp31_S_Probe004	AE008918.1	TGGACGCTCAAGTCGGAATACCTCT	76	25	omp31	Brucella melitensis 16M

Tabla15. Continuación. Conjunto de sondas diseñadas para la detección de agentes susceptibles de ser empleados en bioterrorismo

4.1.4.4 Validación de los ensayos para la detección de patógenos de interés en bioterrorismo mediante ensayos de hibridación

Este proceso se realizó mediante la puesta a punto de las condiciones experimentales necesarias para la realización de los ensayos de hibridación y el análisis de los resultados de las hibridaciones de diferentes muestras frente al conjunto de sondas inmovilizadas sobre la superficie de las micromatrices. La validación de las diferentes sondas fue un proceso secuencial en el que se analizó el comportamiento de las sondas diseñadas frente a cada uno de los organismos de interés y posteriormente se llevaron a cabo ensayos ciegos con muestras problema proporcionadas por grupos colaboradores.

La validación fue realizada en dos etapas. En una primera etapa se analizó la especificidad de las sondas mediante la realización de hibridaciones con muestras que contenían el blanco reconocido por las sondas inmovilizadas sobre la superficie de las micromatrices. En la segunda etapa se analizó la sensibilidad del sistema mediante la titulación de algunas muestras. Estos ensayos de titulación mostraron que la sensibilidad del dispositivo está determinada por la sensibilidad y eficiencia de la reacción de PCR y marcaje pudiendo alcanzarse sensibilidades de hasta 10 copias de la diana de PCR. En la figura 17 se pueden apreciar los resultados de estos ensayos.



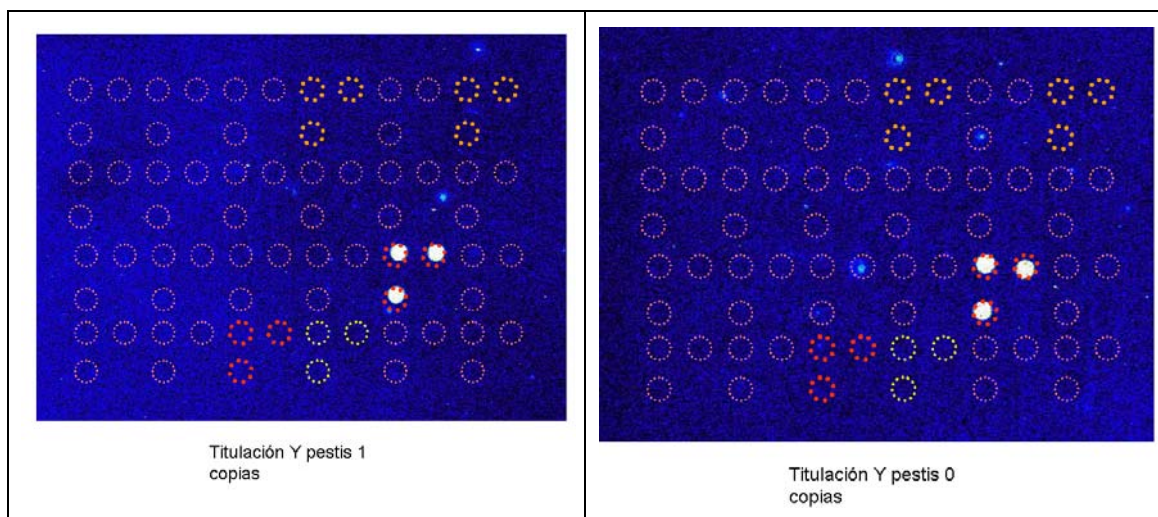
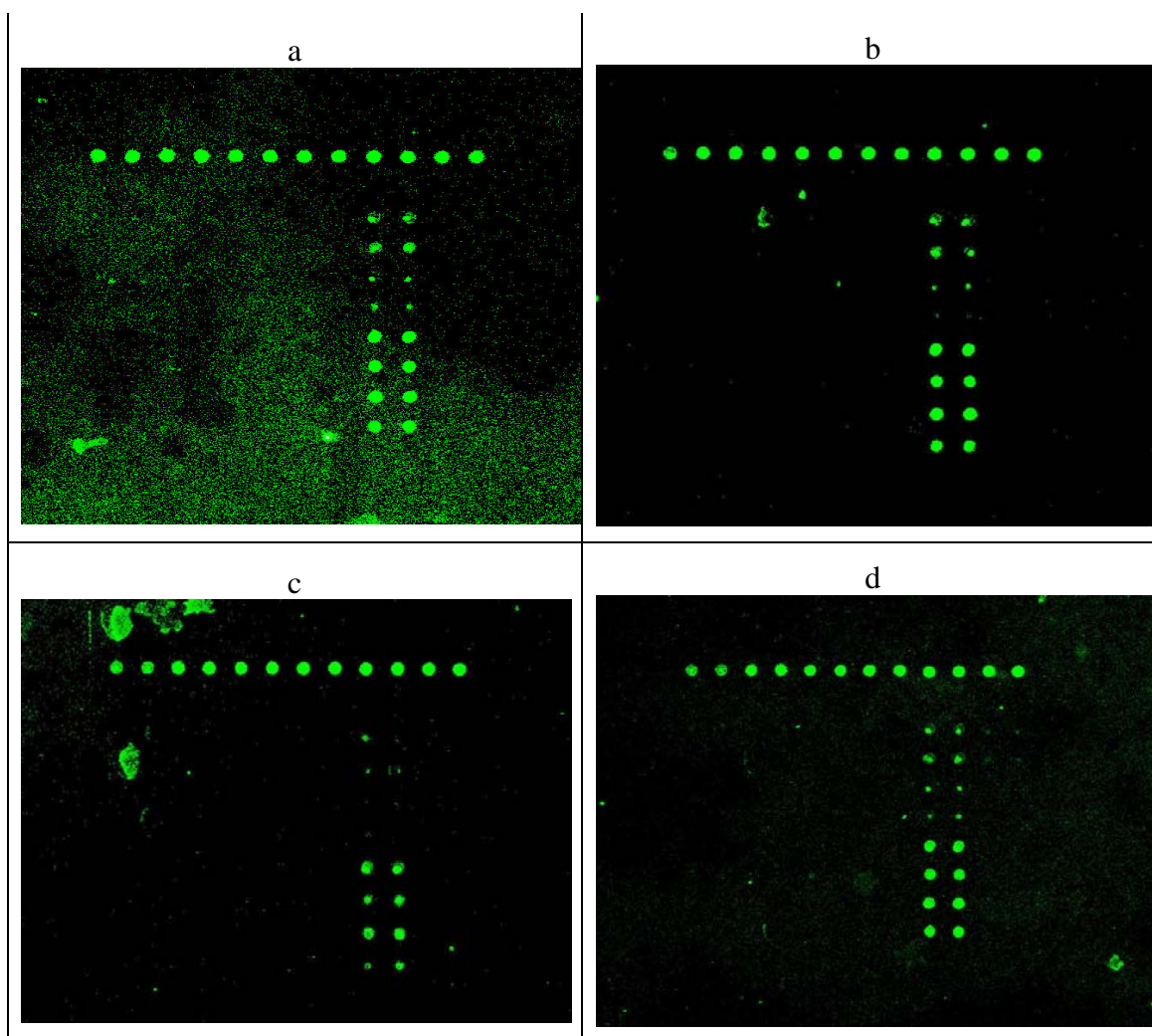


Figura 17a Hibridaciones realizadas empleando diferentes diluciones (5000, 1000, 500, 100, 50, 10, 1 y 0 copias) de plásmido que contiene el gen PLA de *Y. pestis*. Se pueden apreciar las señales procedentes la hibridación positiva con la sonda específica diseñada para la detección de este producto (resaltada con círculos rojos) y del control positivo. Se puede apreciar señal en las hibridaciones empleando diluciones que contienen hasta 10 copias- Por debajo de esa dilución no es posible detectar señal en las hibridaciones.



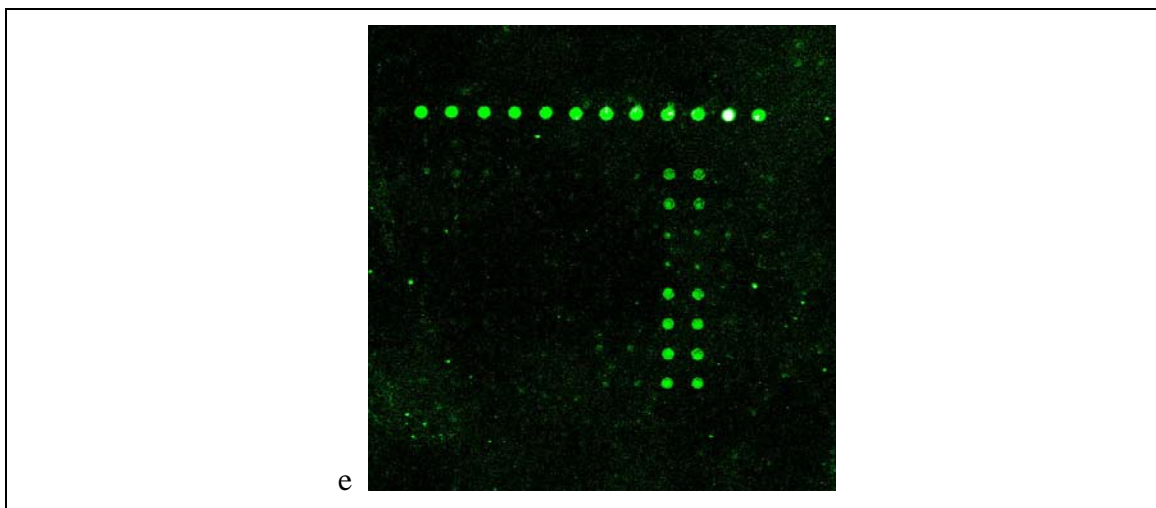


Figura 17b Hibridaciones de Y. pestis 10326 utilizando diferentes concentraciones de material genómico (a, 200pg; b, 20pg; c y d, 2pg; y e, 200fg, equivalente a 78.000, 7800, 780 y 78 copias de genoma) para validar las diferentes sondas diseñadas para la detección de Y. pestis. Además se incluyó una hibridación empleando 100 copias de un plásmido con el gen PLA. En estos ensayos ha sido posible detectar señal positiva en todas las sondas y en todas las diluciones analizadas. Se observa que hay una sonda, “YpestisPlasm:1251U35”, que de manera consistente presenta unos menores niveles de intensidad en todas las hibridaciones y tiene una menor sensibilidad en comparación con el resto de las sondas analizadas

4.1.4.4.1 Validación de las micromatrices empleando cultivos de laboratorio con aguas contaminadas.

Una vez realizada la validación de las sondas a nivel individual empleando muestras control se realizó una segunda etapa de validación de las micromatrices en las que se emplearon muestras procedentes de grupos colaboradores. Se realizaron dos tipos de ensayos, uno en el que se analizaron muestras procedentes de diferentes aguas contaminadas artificialmente con muestras de *F. tularensis* holarctica cepa LVS y muestras ciegas de las que no se conocía su composición hasta una vez realizados los ensayos.

En los ensayos para la detección de *F. tularensis* en aguas contaminadas se emplearon ADNs procedentes de 18 muestras generadas mediante la contaminación experimental de diferentes aguas con *F. tularensis* holarctica LVS y su posterior crecimiento durante diferentes periodos de tiempo a diferentes temperaturas (ver tabla 16).

Los ADNs extraídos de estos cultivos fueron posteriormente amplificados y marcados mediante PCR para la detección del gen *tul4* e hibridados sobre las micromatrices. Los resultados de estas hibridaciones se muestran en la tabla 17. Los resultados de estas hibridaciones muestran la detección y asignación correcta en 17/18 hibridaciones (94.4%), detectándose únicamente un falso negativo en el ensayo.

Muestra	Tipo de agua	Temperatura	Tiempo post-contaminación
5	Agua destilada	37°C	24h
8	Agua destilada	37°C	72h
10	Agua destilada	4°C	6 días
12	Agua de río	-----	0
13	Agua de río	4°C	24h
14	Agua de río	Temp. Ambiente	24h
15	Agua de río	37°C	24h
16	Agua de río	4°C	72h
17	Agua de río	Temp. Ambiente	72h
18	Agua de río	37°C	72h
19	Agua de río	4°C	6 días
20	Agua de río	Temp. Ambiente	6 días
21	Agua de río	37°C	6 días
29	Solución salina	Temp. Ambiente	72h
32	Solución salina	37°C	6 días
39	PBS	37°C	24h
40	PBS	4°C	72h
Control Neg.	-----	-----	-----

Tabla 16. Muestras de aguas contaminadas por *F tularensis* para su análisis sobre las micromatrices de ácidos nucleicos.

Muestra	Reactividad frente a la sonda tul 4	Muestra	Reactividad frente a la sonda tul 4
5	+	18	+
8	+	19	-----
10	+	20	+
12	+	21	+
13	+	29	+
14	+	32	+
15	+	39	+
16	+	40	+
17	+	Control Neg.	-----

Tabla 17 Resultados de las hibridaciones realizadas con las muestras procedentes de aguas contaminadas.

4.1.4.4.2 Validación de las micromatrices empleando muestras problema.

En los ensayos de muestras ciegas se analizaron 16 muestras problema que contenían diversos productos de amplificación marcados fluorescentemente para su detección mediante ensayos de hibridación. En la tablas 18, 19, 20, 21 y 22 se presentan los resultados de las determinaciones realizadas sobre estas muestras problema. Mediante estos ensayos fue posible identificar en un 100% de las muestras la presencia o ausencia de alguno de los agentes de interés. A nivel de las sondas analizadas se identificaron correctamente un 87.5% de las posiciones analizadas.

Muestra	Organismo identificado sobre la micromatriz	Contenido real de la muestra
1	B. anthracis	B. anthracis
2	B. anthracis	B. anthracis
3	B. anthracis	B. anthracis
4	B. anthracis	B. anthracis
5	Negativo	E. coli
6	Brucella	Brucella abortus
7	Brucella	Brucella melitensis
8	Brucella	Brucella suis
9	Francisella	F. tularensis tularensis
10	Francisella	F. tularensis holarctica
11	Francisella	F. tularensis novicida
12	Negativo	Negativo
13	Francisella	F. tularensis tularensis
14	Francisella	F. tularensis holarctica
15	Francisella	F. tularensis novicida
16	Negativo	Negativo

Tabla 18. Resumen de las muestras problema analizadas con los organismos identificados en cada muestra y el contenido real de las muestras. Se puede apreciar que en todos los casos ha sido posible la identificación a nivel genético del organismo presente en la muestra. El estudio realizado empleó la PCR múltiplex de B. anthracis para la detección de los diferentes productos generados durante estas reacciones. Para la detección de Brucella y Francisella se utilizaron las reacciones de amplificación de los genes omp2 y tul4 respectivamente.

Muestra	Sondas analizadas			
	B1-34U	B1-137U	B2-44U	B2-248U
5	----	----	----	----
6	----	----	----	+
7	----	----	----	+
8	----	----	----	+

Tabla 19. Detalle de las reacciones analizadas empleando muestras problema en las que se detectó la presencia de Brucella. Solo se detectaron señales procedentes de una de las sondas diseñadas para la detección del producto de amplificación del gen omp2, en las muestras en las que había DNA de un miembro del género Brucella.

Muestra	Sonda					
	Ba	Cya	Lef3/4	BCA	Lef	Pag
1	+	----	----	----	----	----
2	+	----	----	+	----	----
3	+	----	+	----	+	----
4	+	+	+	----	----	----

Tabla 20. Detalle de las reacciones positivas y negativas observadas en las muestras en las que había DNA de B. anthracis. El signo “+” denota detección positiva y “----” significa no detección. Se puede observar que en todas ellas fue posible la detección del producto Ba813, y en al menos una de las muestras el resto de los productos analizados. Las muestras mezclas de diversos productos de amplificación para los diversos productos en lugar de contener material genómico de B. anthracis contenían los productos de PCR de los fragmentos analizados marcados y listos para su análisis.

Muestra	Sonda					
	Ba	Cya	Lef3/4	BCA	Lef	Pag
1	Presente	Ausente	Ausente	Ausente	Ausente	Ausente
2	Presente	Ausente	Ausente	Presente	Ausente	Ausente
3	Presente	Presente	Presente	Ausente	Ausente	Ausente
4	Presente	Presente	Presente	Presente	Ausente	Ausente

Tabla 21. Contenido real de las muestras 1-4 en las que había productos pertenecientes a *B. anthracis*

Muestra	Sonda					
	Ba	Cya	Lef3/4	BCA	Lef	Pag
1	+	+	+	+	+	+
2	+	+	+	+	+	+
3	+	----	+	+	----	+
4	+	+	+	----	+	+

Tabla 22. Resultados esperados frente a resultados obtenidos en la detección de las muestras con *B. anthracis*. El signo “+” denota coincidencia entre la detección y la presencia del producto en la muestra, “----” significa una discrepancia. La tasa de reconocimientos correctos fue del 87.5% (21/24) con 2 falsos positivos y un falso negativo.

4.1.4.5 Validación de los ensayos para la detección de patógenos de interés en bioterrorismo mediante ensayos de extensión de la sonda inmovilizada.

Los ensayos de extensión de la sonda inmovilizada sobre la superficie de la micromatriz son una alternativa a los ensayos de hibridación. Para la validación de las sondas diseñadas para la realización de estos ensayos se pusieron a punto las condiciones de reacción y se comparó la utilización de diferentes análogos de nucleótido para la detección de las reacciones de extensión.

Un aspecto importante para la realización de estos ensayos fue el diseño de una sonda con capacidad para servir como control positivo dado que forma un bucle que sirve como substrato y molde para su elongación por parte de la polimerasa. Figura 18.

Nombre	Longitud	Tm	ΔG	Secuencia
Ctrl_self_PrimerExt	80	89.4	-29.8	ATGGTAGCCGCTTATGCTGGAATG GTTCTGTCATATACTGATGAGGCA TTGTCATCAAAAGATGACAATGCC TCATCAGT

Figura 18 Secuencia y características de la sonda diseñada como control interno para los ensayos de extensión de primer, en azul está resaltada la zona que sirve como cebador para los ensayos de extensión de primer. Imagen de la estructura secundaria que se forma cuando se emplea

Para la realización de los ensayos de extensión se emplearon dos aproximaciones diferentes, en una primera aproximación se analizó la extensión de las diferentes sondas sobre la superficie de la micromatriz empleando dNTPs en los que uno de ellos es sustituido por un análogo de nucleótido marcado fluorescentemente, siendo posible detectar la extensión de las sondas para los productos presentes en las muestras. Figura 19

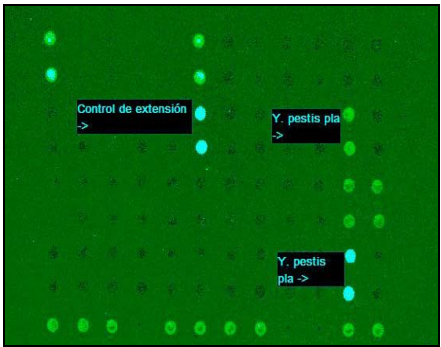


Figura 19. Ejemplo de un ensayo de extensión de primer inmovilizado sobre la superficie empleando una muestra que contiene *Y. pestis*, *F. tularensis tularensis* (tul4), una pequeña cantidad de *B. melitensis melitensis* (bcsp31) y *B. mallei* (gyrA). Se puede apreciar que en algunos puntos se ha alcanzado la saturación de la señal (puntos blancos en la imagen)

La segunda aproximación seguida para la realización de los ensayos consistió en la extensión de las sondas empleando los cuatro ddNTPs, estando cada uno de ellos marcado

con un fluoróforo diferente para su detección. Esta otra aproximación también permite la detección de los diferentes productos presentes en las muestras analizadas pero además aporta información acerca de la base incorporada en cada momento. En la figura 20 se muestran los resultados de estos ensayos de validación.

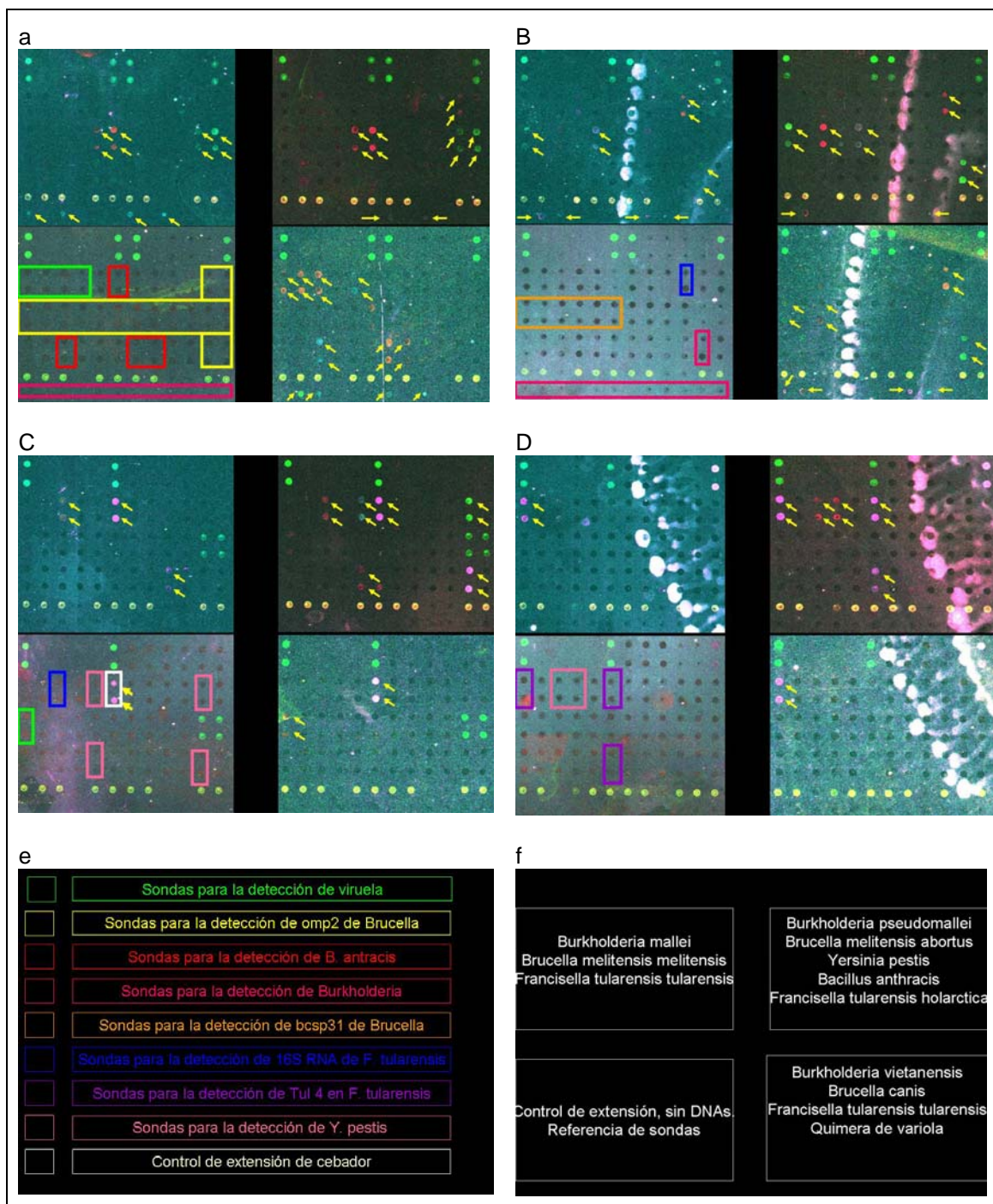


Figura 20. Resultados de reacciones de extensión de primer empleando ddNTPs empleando muestras procedentes de diversos organismos. Las imágenes(A-D) contienen los resultados de 3 mezclas de muestras y una muestra control negativo sin ADN y están construidas mediante la superposición de las señales procedentes de los cuatro canales analizados (uno para cada ddNTP, rojo Cy5-ddCTP, verde Cy3-ddUTP, rosa TexasRed-ddATP y azul 5FAM-ddGTP).y en ellas se puede apreciar la incorporación de los diferentes ddNTPs y diferencias entre las muestras dependiendo de las sondas reconocidas durante el ensayo.

4.1.5 Detección de adenovirus respiratorios

4.1.5.1 Diseño de sondas para la detección de adenovirus humanos implicados en patologías respiratorias

Se diseñaron sondas para la detección específica de los serotipos de adenovirus más importantes en las infecciones respiratorias en humanos, serotipos 1, 2, 3, 4, 5, 7, 14 y 21. Estas sondas fueron diseñadas para identificar los diferentes serotipos en base a las diferencias existentes en la región codificante de la proteína hexon definida por la reacción de amplificación (Coiras *et al.*, 2003). Para la delimitación de las secuencias sobre las que se diseñaron las sondas se emplearon las secuencias disponibles en la base de datos Genbank en el año 2004 y procesadas mediante la herramienta BUSSUB, que se describe más adelante. Los oligonucleótidos seleccionados como sondas durante este proceso se presentan en la tabla 24.

Probes	Sequence	Length	Tm (°C)
gADV	TTYAACCAYCACCGYAAYGCKGG	23	77.3
ADV-a	TTGCTGGGCAACGGTCGCTACGTGC	25	83.9
ADV-b	TGTTGTTGGGAAACGGCCGC	19	75.1
ADV-c	GCTGGTCTGCGCTACAGGTCCATGC	25	83.9
ADV-d	GCTACGTGCCATTCCACATCCAGGT	25	80.6
ADV-e	TCTCCTGCCGGGCTCATACACCTAC	25	82.3
ADV-f	CTGGYCTACGCTRCCGGTCCATGCT	25	82.3
ADV-g	CCGCCATCTCTAGTAGACACCTATG	25	79
ADV-h	GCTTTTGGGCAATGGTCGTTACGTG	25	79

Tabla 24 Sondas diseñadas para la detección de los principales adenovirus causantes de patologías respiratorias.

4.1.5.2 Validación de los ensayos para la detección de virus respiratorios mediante el análisis de muestras clínicas.

El desarrollo de un ensayo para la validación de las micromatrices para la identificación de agentes infecciosos en muestras clínicas empleando micromatrices de ácidos nucleicos fue diseñado para la detección de los principales adenovirus causantes de patología respiratoria en humanos.

La micromatriz que se diseñó para llevar a cabo estos ensayos contenía una sonda genérica para la detección de cualquier adenovirus presente en la muestra y sondas para la detección simultánea de los principales serotipos de adenovirus involucrados en patologías respiratorias en humanos. Estas sondas fueron diseñadas para generar un patrón de hibridación diferente para cada uno de los serotipos analizados y fueron validadas usando muestras cultivadas en laboratorio de los diferentes serotipos de interés. Figura 21. Estos patrones característicos e individuales para la detección de cada virus permitieron el reconocimiento del serotipo de los adenovirus presentes en las muestras mediante una sencilla inspección visual.

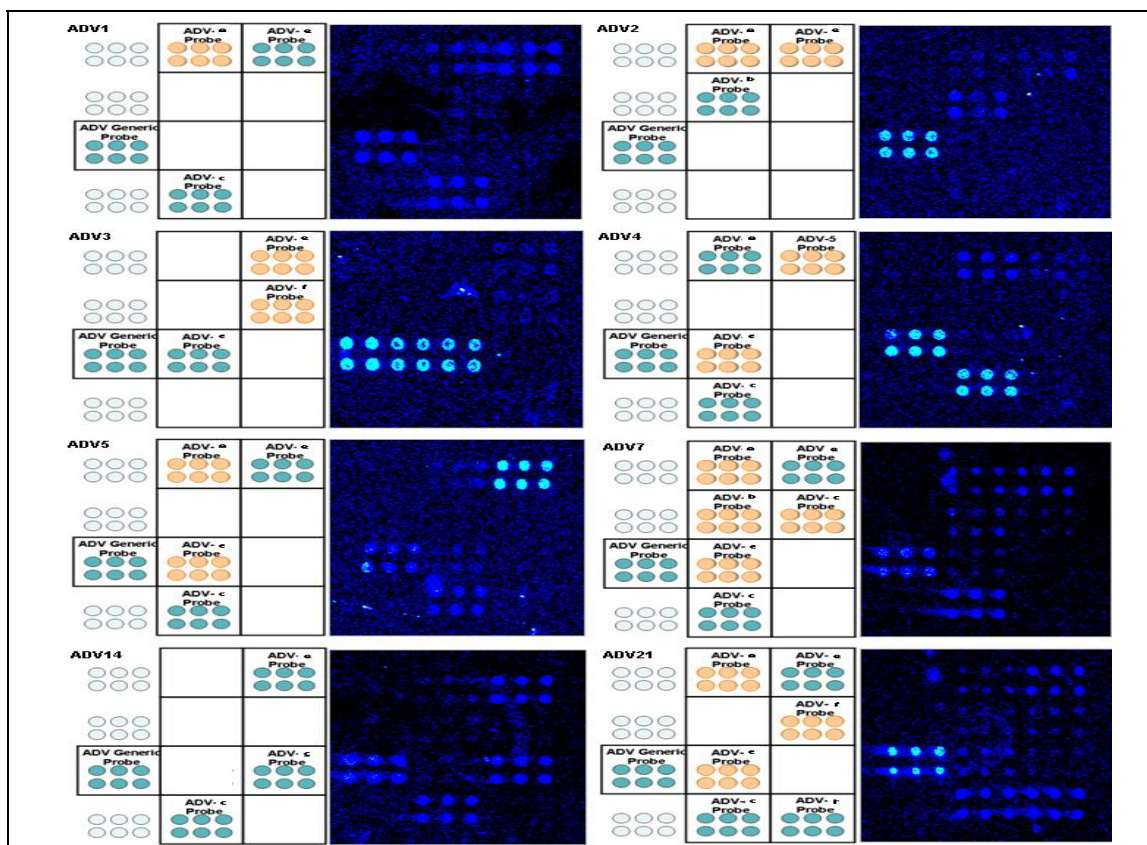


Figura 21 Patrón de detección de los principales adenovirus respiratorios al reconocer diferentes sondas con diferentes intensidades

La especificidad de las sondas y los patrones diseñados fueron ensayados frente a un panel consistente en los genomas completos de los serotipos (1-49) de referencia de los adenovirus, no encontrándose ninguna reactividad cruzada significativa distinta de las esperadas para los serotipos 1, 2, 3, 4, 5, 7, 14, y 21. Figura 22.

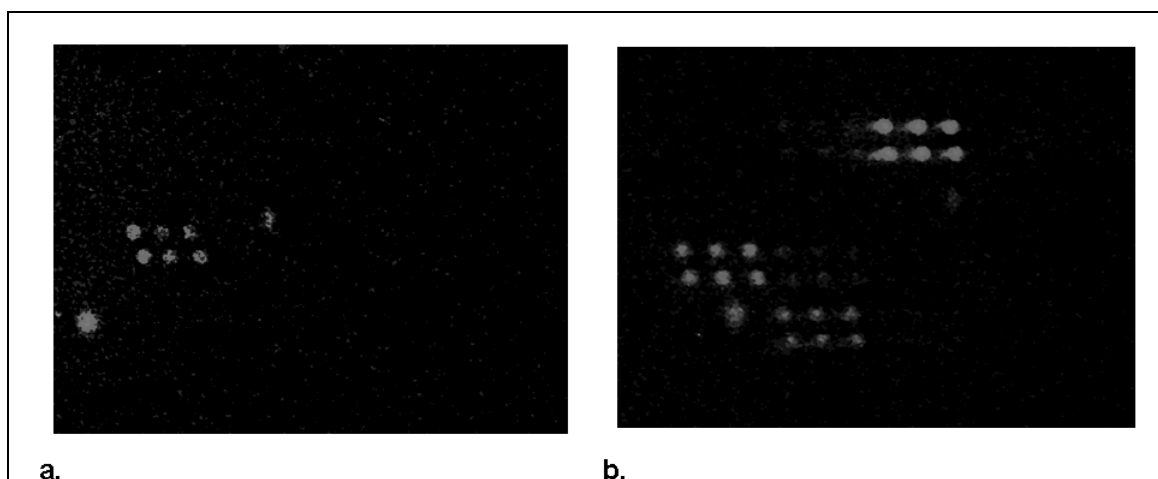


Figura 23 Imagen con la hibridación de dos muestras de adenovirus marcadas. En la figura a la muestra hibridada es una adenovirus 19 y solo se puede apreciar señal procedente de la sonda genérica para la detección de los adenovirus. En la figura B se puede apreciar la señal procedente de una hibridación con el adenovirus 5 y el patrón de señales de este adenovirus.

Además de realizar los ensayos para analizar la reactividad cruzada frente a otros adenovirus, se analizaron las posibles interacciones con otros virus respiratorios que pudieran

estar presentes en las muestras, no observándose ninguna interacción con muestras consistentes en virus influenza de los tipos A, B y C y el virus respiratorio sincitial de tipo A y B.

Los ensayos de sensibilidad para los ensayos sobre las micromatrices se realizaron empleando controles con productos de amplificación del adenovirus de subtipo 1 clonados y que contenían 10, 100 y 1000 copias del producto de amplificación. Se identificó un umbral de detección de entre 10 y 100 copias para la detección genérica de los adenovirus en la muestra y de entre 100 y 1000 copias para la identificación de los serotipos empleando las cepas prototipo.

Los ensayos para validar el dispositivo sobre muestras clínicas se enfocaron en el genotipado y asignación a uno de los serotipos anteriormente descritos. Para realizar estos ensayos, se emplearon 10 muestras procedentes de aspirados nasofaríngeos que fueron clasificados como infecciones adenovirales por PCR, IFI o aislamiento en cultivos celulares que fueron analizadas empleando las micromatrices desarrolladas. Una vez realizados los experimentos, las micromatrices fueron escaneadas. La asignación de cada una de las imágenes a cada uno de los serotipos fue realizada inicialmente mediante la comparación visual con los patrones de hibridación anteriormente descritos (ver figura 21).

Una segunda fase del análisis consistió en el análisis numérico de las intensidades cuantificadas para cada uno de los puntos analizados y la asignación a uno de los patrones generados por los diferentes genotipos respiratorios. Figura 24.

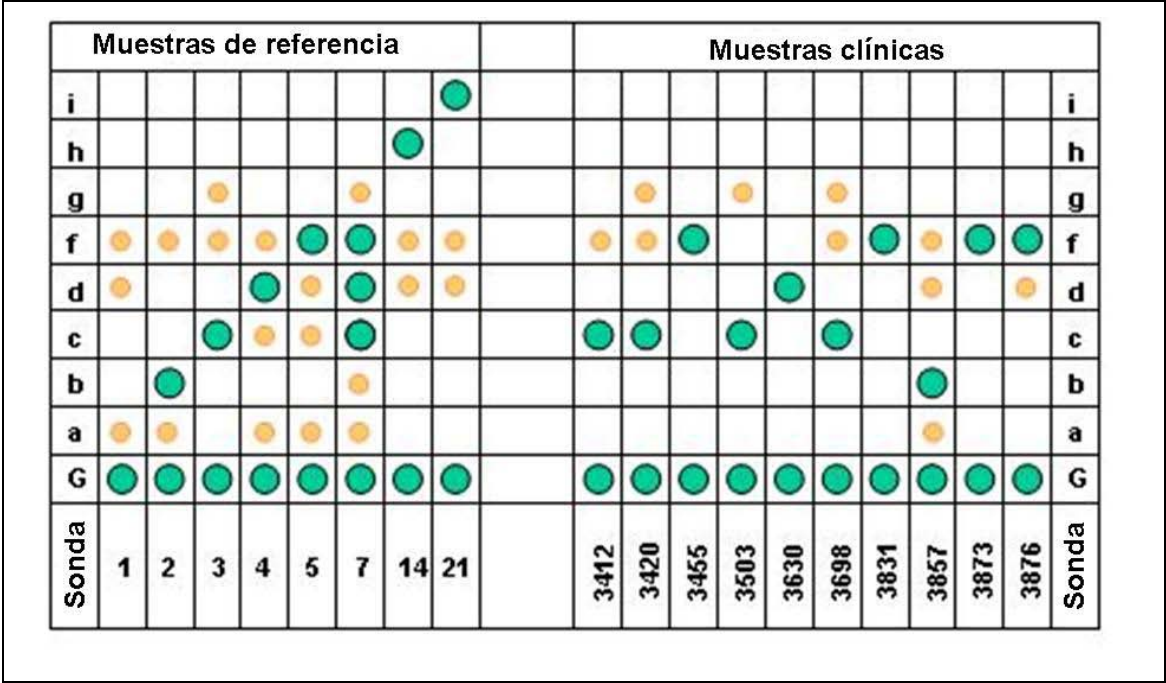


Figura 24. Imagen generada con las intensidades medidas de los puntos considerados positivos para las muestras control y para las muestras clínicas analizadas. El diámetro de los puntos representa las diferentes intensidades detectadas durante las hibridaciones. Los puntos verdes grandes representan mayores intensidades que los puntos naranja más pequeños.

Los adenovirus detectados en estas muestras fueron amplificados y secuenciados para realizar un análisis filogenético y su comparación con los resultados y la asignación a los

diferentes serotipos realizados sobre las micromatrices. El estudio filogenético de estas muestras se realizó empleando un método de inferencia basado en el algoritmo de neighbor-joining y 10000 réplicas mediante técnicas de bootstrap. Figura 25

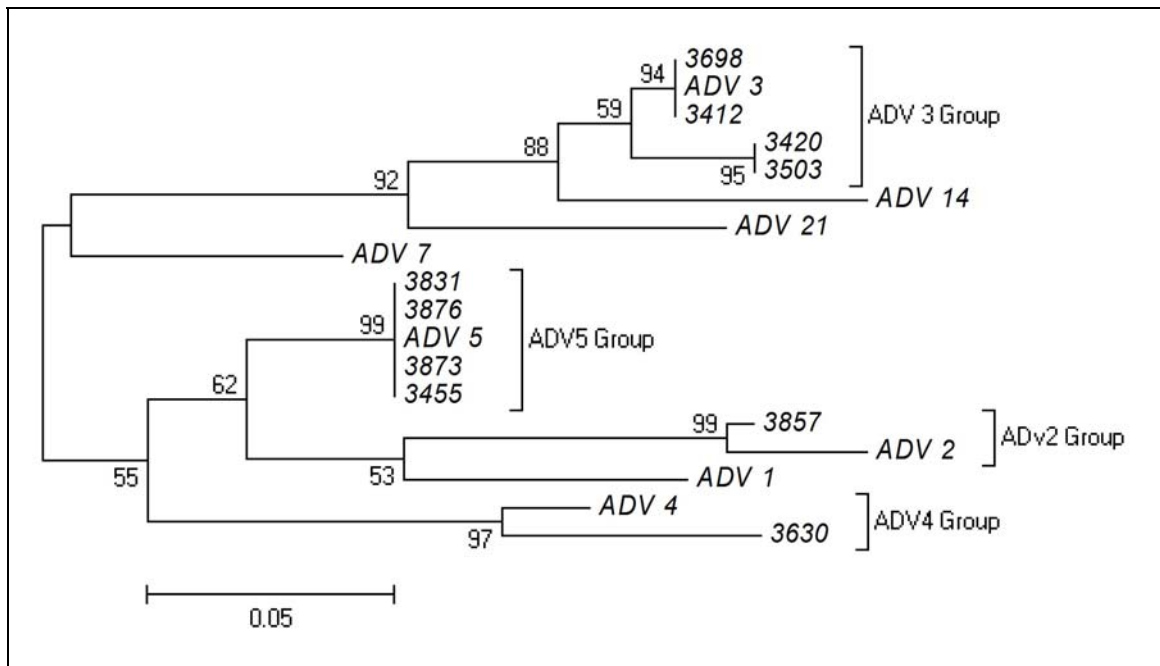


Figura 25 Análisis filogenético de las muestras clínicas empleadas durante los ensayos de validación de las micromatrices. Las secuencias de referencia de los diferentes adenovirus empleados para asignar el genotipo de cada una de las muestras clínicas fueron obtenidas en Genbank.

Los resultados de los ensayos realizados mostraron en todos los casos (100%) coincidencia entre las asignaciones a los diferentes serotipos realizadas mediante los ensayos con micromatrices con los resultados obtenidos mediante la secuenciación y análisis filogenético. En paralelo se emplearon también ensayos en los que se emplearon otras muestras clínicas infectadas con otros virus respiratorios, como influenza o RSV, como controles de la especificidad de los ensayos. En ninguno de estos ensayos se apreció ningún tipo de señal cuando el virus presente en las muestras era distinto de un adenovirus.

4.2. Análisis cuantitativo de la respuesta inmune humoral frente a la infección por diferentes subtipos de VIH empleando micromatrices de material biológico:

El análisis de datos de micromatrices es una etapa crucial y clave para la obtención de resultados y conclusiones en este tipo de ensayos. La mayor parte de los trabajos que se pueden encontrar en la literatura relativos al análisis de datos se han enfocado el análisis de datos de expresión génica. En el caso de los ensayos con micromatrices de proteínas estos análisis presentan algunas particularidades que provocan la necesidad de desarrollar o adaptar las técnicas y estrategias existentes para el análisis de la expresión génica a nivel de ARNm.

4.2.1 Cuantificación de las imágenes, control de calidad y normalización de los datos (preprocesamiento y transformación)

La primera etapa del análisis de los datos de los ensayos con micromatrices de material biológico consiste en la cuantificación de las imágenes generadas durante los ensayos de afinidad. Esta cuantificación consiste en la extracción de valores numéricos de intensidad de la señal a partir de las imágenes. Posteriormente estos valores numéricos son los que se van a emplear en el resto de las etapas de análisis de datos.

4.2.1.1 Cuantificación de las imágenes

Las micromatrices, tras ser incubadas para la reacción antígeno-anticuerpo, fueron escaneadas en un escáner láser para la generación de las imágenes características de este tipo de ensayos. Una vez generadas estas imágenes fueron cuantificadas usando el programa bioinformático ImaGene 4.0 (Biodiscovery. Inc. Ca. USA). El proceso de “gridding” o localización espacial de las posiciones impresas sobre la superficie de la micromatriz se realizó de manera manual e individualizada para cada una de las imágenes analizadas (ver figura 26).

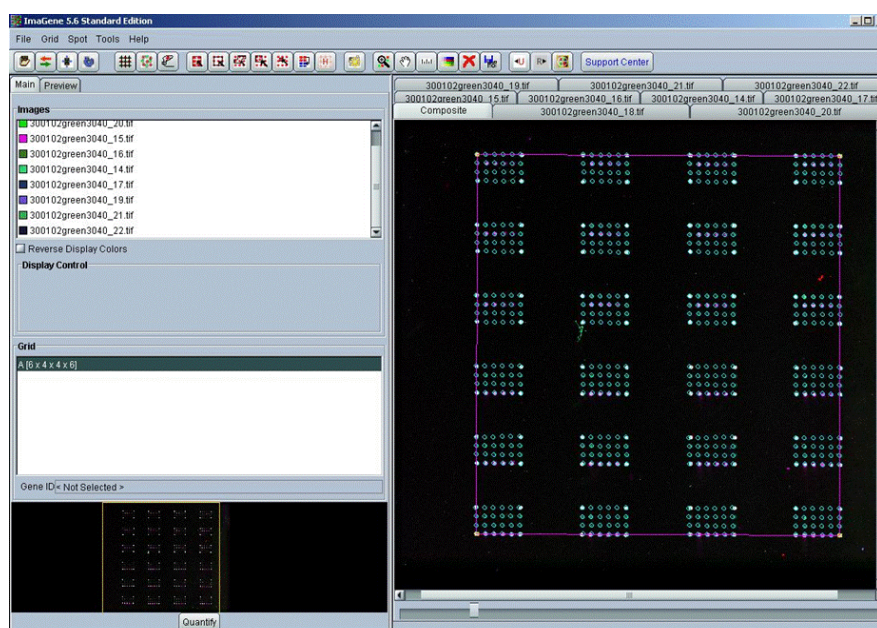


Figura 26 Figura en la que ejemplifica el proceso de “gridding” de la primera fase del análisis de datos.

Una vez realizado este proceso de alineamiento de la matriz teórica sobre la imagen se procedió a la segmentación de las imágenes para la determinación de los valores de señal y ruido de fondo para cada una de las posiciones impresas. El programa empleado proporciona diferentes opciones para el proceso de segmentación de las imágenes incluyendo diferentes algoritmos de segmentación. La selección del algoritmo y parámetros de segmentación se llevó a cabo tras analizar dos de las imágenes y comparar los resultados de las mismas (resultados no mostrados), seleccionándose el algoritmo de segmentación basado en histogramas con los parámetros proporcionados por defecto para este algoritmo.

El resultado de este proceso de segmentación y cuantificación de las imágenes consistió en un archivo para cada una de las imágenes en el cual se reflejan diferentes parámetros tales como señal, ruido de fondo, uniformidad de los puntos y diferentes valores estadísticos asociados con la determinación de los parámetros anteriormente reseñados. Estos valores estadísticos son empleados para los análisis de control de calidad y para la determinación de los valores de positividad de las reacciones para cada uno de los puntos impresos sobre la superficie de la micromatriz.

4.2.1.2 Preprocesamiento de los datos – Control de calidad

El primer paso del proceso de análisis se ha enfocado como una etapa de control de calidad. Un aspecto fundamental en el trabajo con las micromatrices de material biológico es el control de calidad de los resultados y la reproducibilidad de los resultados. En las micromatrices analizadas durante este trabajo se incluyeron un gran número de réplicas de los diferentes puntos. Cada micromatriz estaba constituida por un conjunto de submatrices que contenían todas las sondas analizadas y que estaban replicadas 12 veces a lo largo de la superficie de la matriz. Para determinar la reproducibilidad de los resultados obtenidos en los ensayos se emplearon técnicas basadas en el análisis ANOVA con $p=0.05$ para descartar la existencia de diferencias estadísticamente significativas entre los valores cuantificados de las distintas submatrices. Aquellos experimentos que no superaron este análisis fueron descartados y no se incluyeron en fases posteriores de análisis.

4.2.1.3 Preprocesamiento de datos - Normalización

El análisis simultáneo sobre la superficie de las micromatrices de las reacciones de un mismo suero frente a los diferentes antígenos inmovilizados permite la realización de ensayos de comparación de reactividad entre sueros frente a estos antígenos.

Un aspecto fundamental para posibilitar la comparación de los datos procedentes de los ensayos de distintas micromatrices es la normalización de los resultados obtenidos de los procesos de cuantificación de las imágenes. La normalización propuesta para el análisis de este tipo de ensayos es una normalización basada en la utilización de los valores de la IgG 90 para referenciar el resto de los valores al promedio de los valores de intensidad observados para la IgG 90. De esta manera se obtienen unas matrices de datos normalizadas que van a servir como sustrato para el resto de los análisis y para una segunda etapa de control de calidad. Durante los ensayos se incluyeron algunas réplicas técnicas (replicas en las que se reacciona el mismo suero sobre diferentes micromatrices) a las que se aplicó de nuevo un análisis ANOVA ($p=0.05$) sobre los datos normalizados para descartar que hubiera diferencias estadísticamente significativas entre los diferentes ensayos.

En el análisis de los datos de expresión génica los datos suelen sufrir una transformación a escala logarítmica. En el caso de los datos procedentes de los ensayos con proteínas se ha analizado el efecto de este tipo de transformaciones. La valoración de la utilidad de este tipo de transformación se ha efectuado analizando el comportamiento de las replicas estudiadas

previamente en el análisis ANOVA ($p=0.05$), observándose que la transformación de los datos permite que algunas réplicas técnicas que anteriormente no habían superado el test ANOVA lo superen.

La utilización durante el diseño experimental de diluciones de los péptidos de interés ha permitido el desarrollo de otro criterio de control de calidad basado en el análisis de los ajustes de regresión lineal de las intensidades medias (ver figura 27). Este tipo de ajuste y de control de calidad permite en algunos casos identificar valores anormales de los datos y permite su eliminación para las siguientes etapas de los análisis.

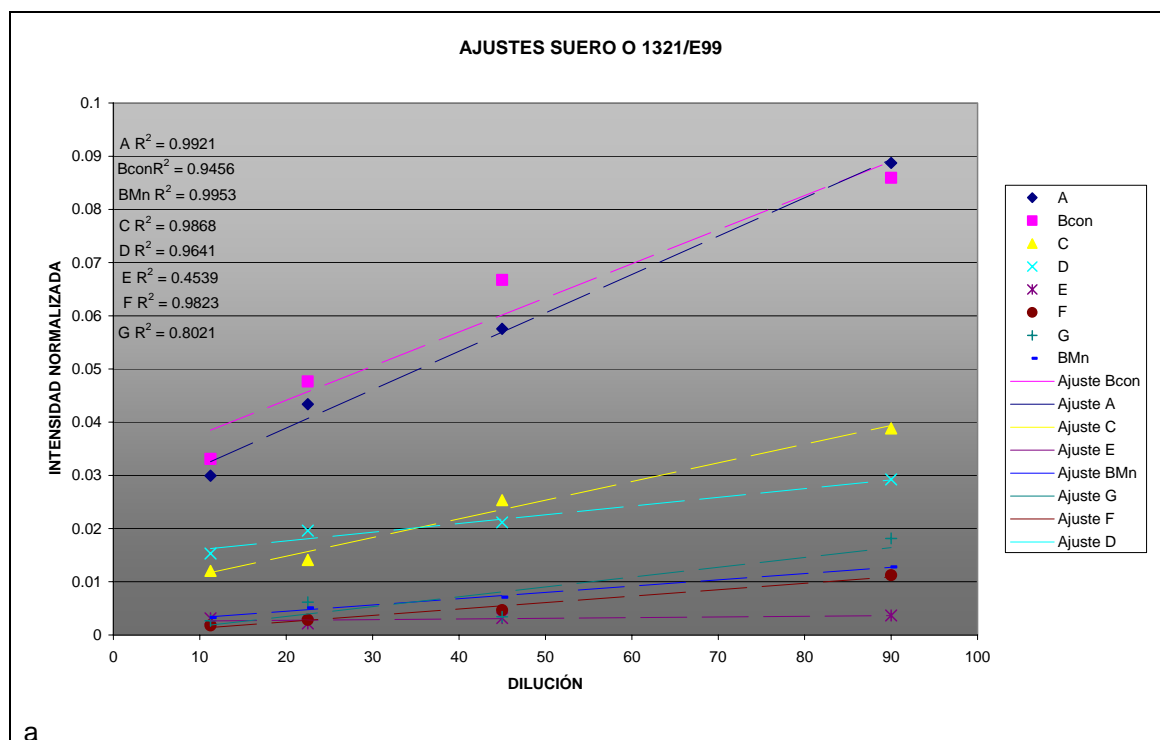


Figura 27 a. Ajustes de las rectas generadas al analizar la respuesta frente a las diferentes diluciones de los péptidos inmovilizados sobre la superficie de las micromatrices de un suero de subtipo O.

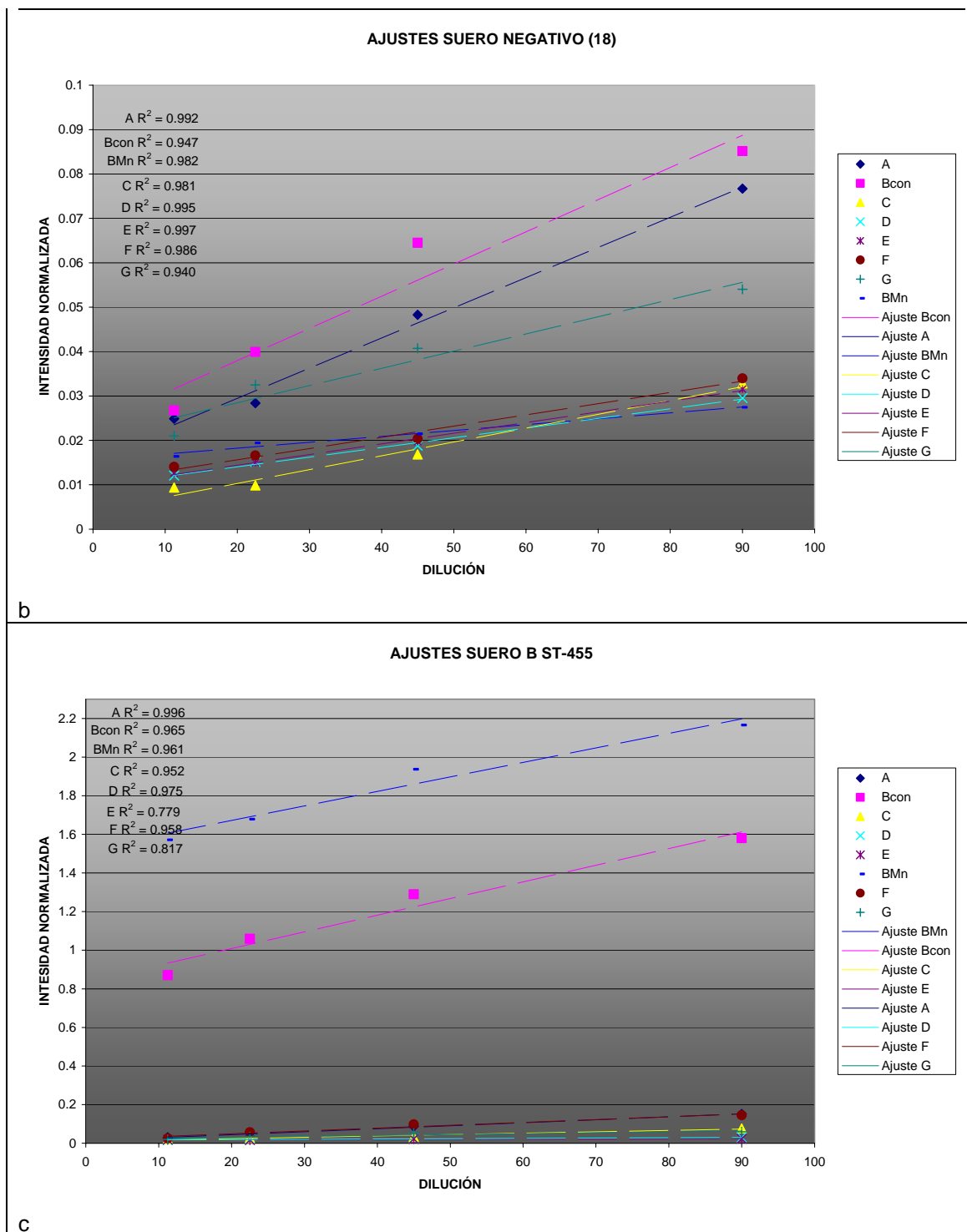
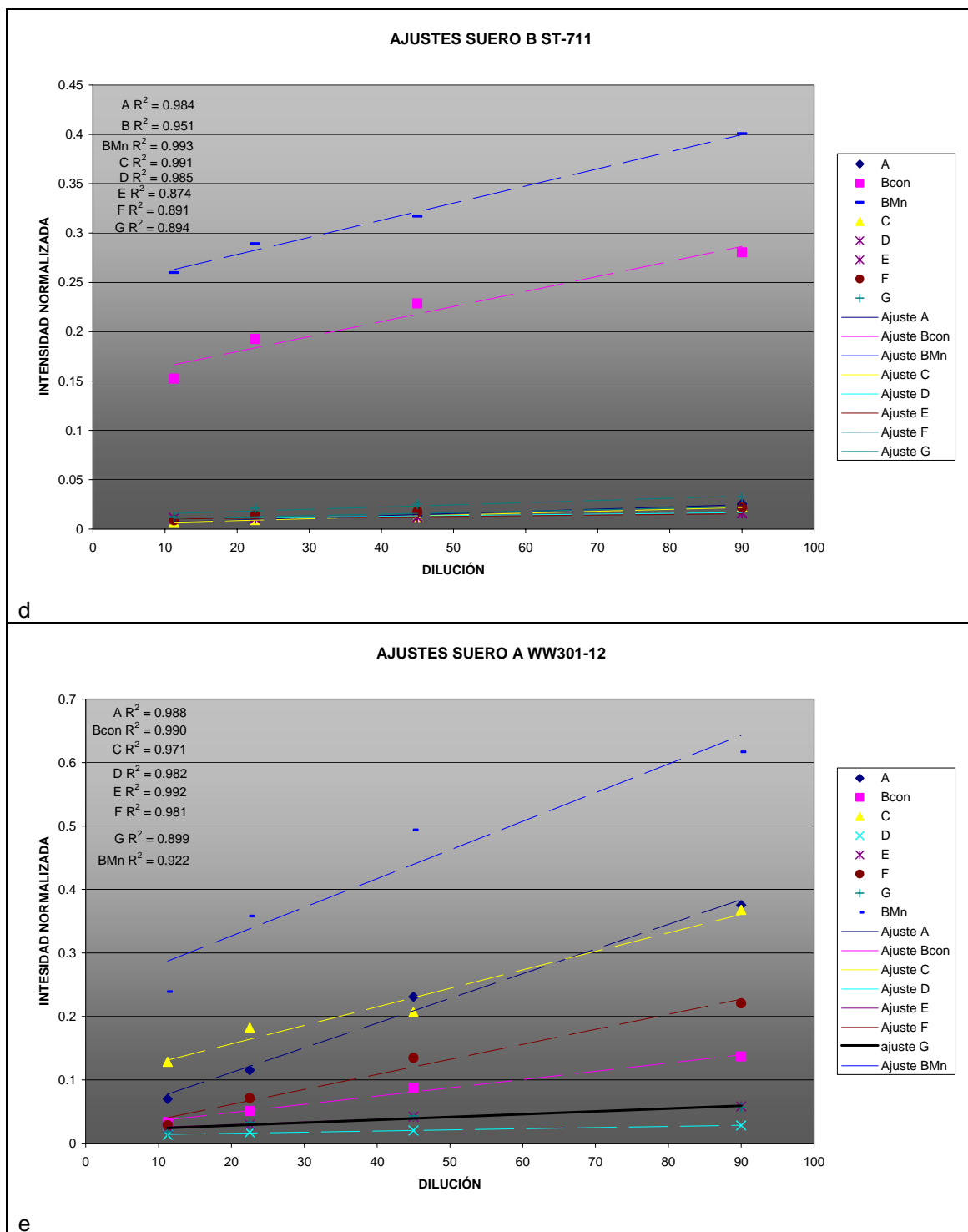


Figura 27 (continuación). Ajustes de las rectas generadas al analizar la respuesta frente a las diferentes diluciones de los péptidos inmovilizados sobre la superficie de las micromatrices de un suero negativo (27b) y un suero de subtipo B (27c)



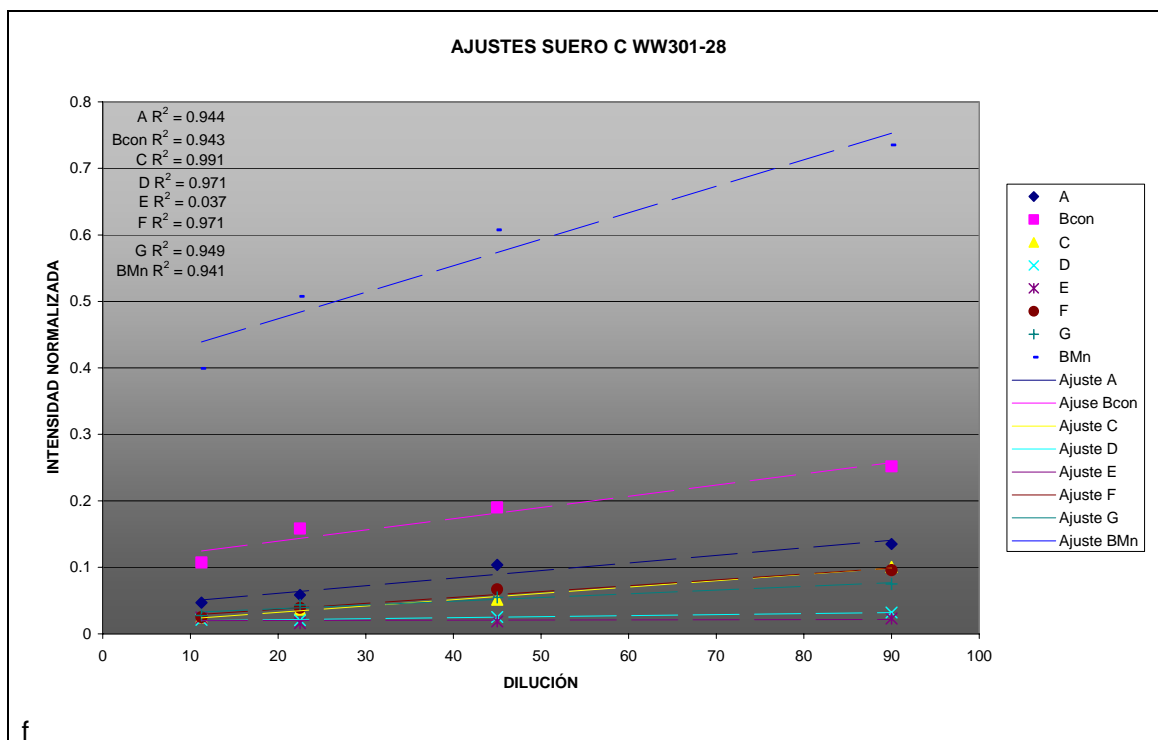


Figura 27 (continuación). Ajustes de las rectas generadas al analizar la respuesta frente a las diferentes diluciones de los péptidos inmovilizados sobre la superficie de las micromatrices de un suero débi de subtipo C (27f)

4.2.2 Determinación de positividad de las reacciones:

A continuación y una vez se seleccionaron aquellos experimentos que superaban el control de calidad determinado por los test ANOVA, la siguiente etapa consistió en la determinación de un criterio de positividad de las reacciones sobre la superficie de la micromatriz. Para determinar la positividad de las reacciones se siguieron dos criterios. Inicialmente se empleó una apreciación subjetiva para determinar aquellas posiciones en las que visualmente se podían apreciar puntos con señal lo suficientemente intensa como para resultar positivos. El segundo criterio que se empleó fue un criterio basado en el análisis numérico de las imágenes cuantificadas, consistió en la determinación de un criterio o valor de corte (cut-off) generado a partir de los datos experimentales observados para todos los ensayos. Para la selección de este valor de corte, se empleó el siguiente criterio:

Si Señal > Ruido + 3 Desv. Est. Ruido → Positivo

Si Señal < Ruido + 3 Desv. Est. Ruido → Negativo

Siguiendo este criterio se analizaron los resultados de las micromatrices con la finalidad de determinar aquellas posiciones en las que hay reacciones positivas. La determinación de estos valores y criterios de positividad son empleados posteriormente a la hora de clasificar los sueros analizados como positivos o negativos.

4.2.3 Comparación entre sueros positivos, negativos y comparaciones entre subtipos en los ensayos con micromatrices:

La primera etapa de este análisis comparativo de la reactividad de los diferentes sueros consistió en la realización de varios diagramas de escarcha o "scatterplots". Estos diagramas representan de una manera sencilla e intuitiva las diferencias existentes en la reactividad entre las categorías analizadas. Mediante esta técnica es posible identificar en las comparaciones un grupo de péptidos cuya diferente intensidad puede servir para diferenciar las categorías analizadas. Para la realización de estos diagramas se agruparon los sueros analizados en base a sus diversas características biológicas y se promediaron en función de estas agrupaciones los valores de intensidad obtenidos frente a cada péptido. Las comparaciones que se llevaron a cabo en esta etapa del análisis fueron las que se detallan a continuación:

- Sueros positivos frente a sueros negativos
- Sueros de tipo M frente a sueros de tipo O
- Comparaciones de los sueros de los diferentes subtipos contra los sueros negativos.

En el análisis comparativo de las respuestas de los sueros positivos frente a los sueros negativos se puede apreciar en los sueros negativos la reactividad que es menor que la que se observa en los sueros infectados, además también se puede apreciar la linealidad de la intensidad de la señal respecto a las diluciones de los sueros analizados. En esta gráfica se destacan por su mayor intensidad las reacciones frente a los péptidos B consenso y Bmn, que son las que presentan una mayor diferencia de intensidad respecto a los sueros negativos. (Figura 28)

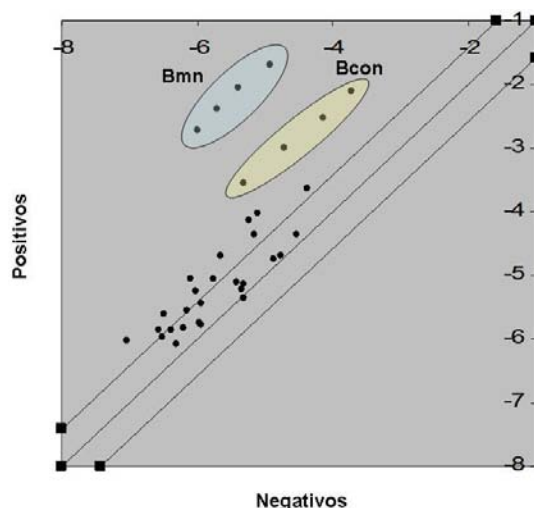


Figura 28 Scatter Plot que representa los valores de las intensidades medias para cada péptido de los sueros positivos frente los valores de los sueros negativos

El resultado de los análisis empleando los scatterplots de los sueros de tipo M frente a los sueros negativos es similar al obtenido al analizar todos los sueros positivos, destacando en la gráfica los puntos que representan los péptidos Bcon, Bmn, C y F. (Figura 29).

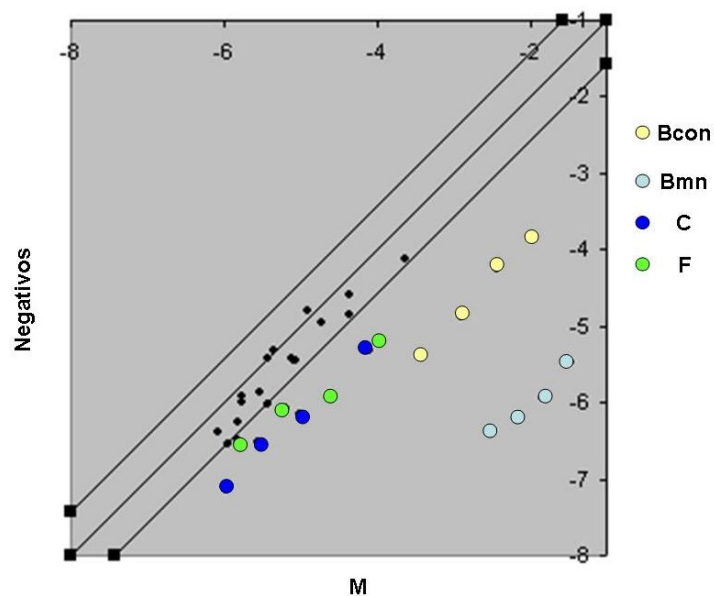


Figura 29 Scatter Plot que representa los valores de las intensidades medias para cada péptido de los sueros positivos de tipo M frente los valores de los sueros negativos

En el análisis de los dos sueros de tipo O frente a los sueros negativos se aprecia que los únicos péptidos frente a los que existe una reactividad diferente son los sueros específicos de tipo O, siendo la reactividad frente al resto de los péptidos muy similar entre los sueros negativos a VIH y los sueros del tipo O (Figura 30).

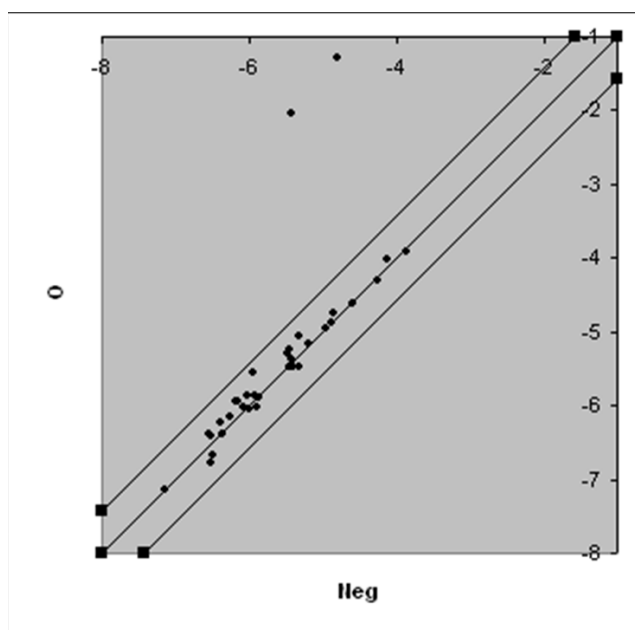


Figura 30 Scatter Plot que representa los valores de las intensidades medias para cada péptido de los sueros positivos de tipo O frente los valores de los sueros negativos, y en el que solo se aprecian como con una intensidad diferente las señales de las 2 diluciones del péptido O

La comparación empleando los scatterplots entre los sueros de tipo M y los de tipo O arrojan unos resultados en los que destacan las grandes diferencias en reactividad frente a los péptidos O, Bcon, Bmn, C y F. Estos resultados son similares a los que surgirían de intersectar las gráficas resultantes de los dos ensayos anteriores. (Ver figura 31).

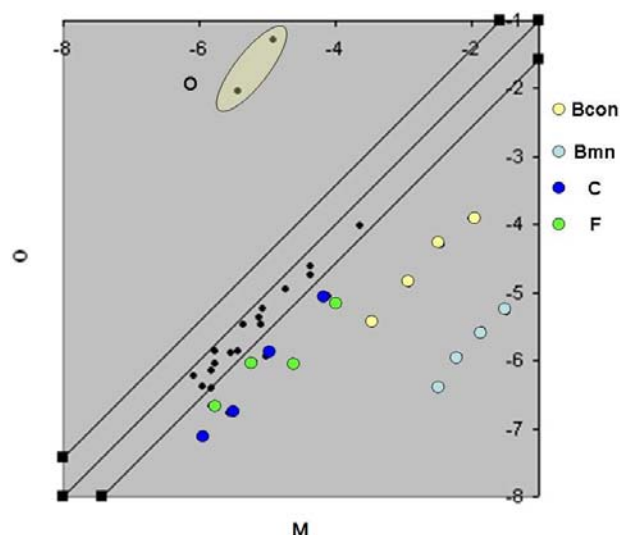


Figura 31 Scatter Plot que representa los valores de las intensidades medias para cada péptido de los sueros positivos de tipo M frente los valores de los sueros positivos de tipo O

Las comparaciones de reactividad entre los diferentes subtipos y los sueros negativos están representadas en la figura 32 y los péptidos más destacados en cada una de estas comparaciones están recogidos en la tabla 25. Estos resultados muestran la existencia de unos “perfiles de reactividad” característicos para cada subtipo que los diferencian de los sueros negativos, a excepción de los subtipos B y C que a la luz de estos resultados compartirían los péptidos característicos pero con unos perfiles muy diferentes tal y como se puede observar en la figura.

Como se ha indicado anteriormente los resultados de los análisis empleando “scatterplots” parecen indicar la existencia de “perfiles de reactividad” característicos para cada uno de los subtipos. Por ello, en una siguiente etapa del análisis de resultados, se ha recurrido a la aplicación de técnicas numéricas de comparación de clases que originalmente han sido desarrolladas para el análisis de datos de expresión génica sobre micromatrices. Se han empleado técnicas de análisis supervisadas que han permitido la comparación de los valores de intensidad observados durante los ensayos con la finalidad de identificar cuales de los elementos impresos sobre la superficie de las micromatrices sirven para diferenciar de una manera estadísticamente significativa los diferentes subtipos analizados.

Subtipo	Péptidos Representativos	Subtipo	Péptidos Representativos
A	A, Bcon, Bmn, C, E y F	D	A, Bcon, Bmn, D y F
B	Bcon, Bmn, C y F	E	A, Bmn, C, E y F
C	Bcon, Bmn, C y F	F	Bcon, Bmn y F

Tabla 25 Péptidos identificados como discriminantes en los ensayos de scatterplot para la discriminación entre los diferentes subtipos

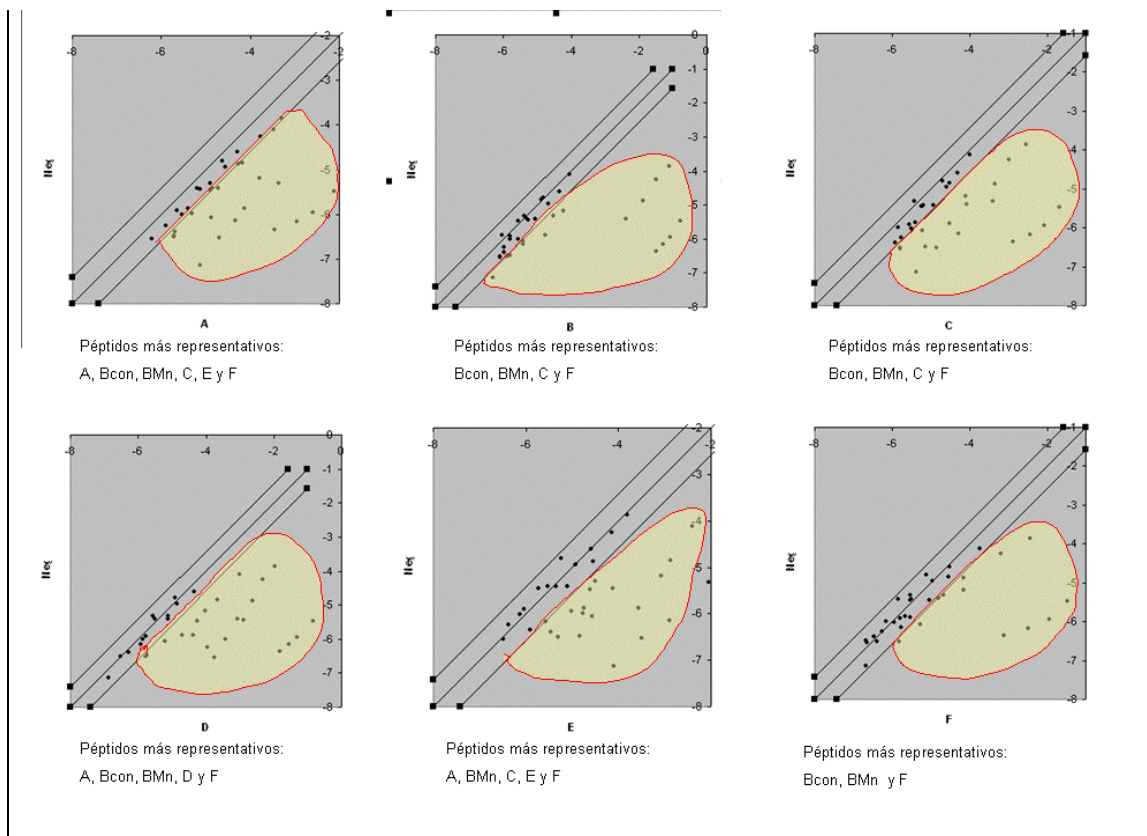


Figura 32 Scatter Plot que representa los valores de las intensidades medias para cada péptido de los sueros positivos de los diferentes subtipos analizados frente a los sueros negativos

Los primeros ensayos que se han realizado han estado dirigidos a la identificación de los péptidos que resultan más significativos para la determinar la positividad de un suero frente a antígenos de VIH. Para ejecutar estos ensayos se han utilizado tres aproximaciones diferentes, empleando únicamente los sueros de tipo N, excluyendo de la misma los dos sueros de tipo O presentes en el análisis. En una primera aproximación se limita el número máximo de posibles falsos positivos con una confianza del 99% a 1 péptido identificado como un posible falso positivo. Las otras dos aproximaciones consisten en la aplicación de la misma técnica, un test multivariante con 1000 permutaciones, con dos niveles de significancia diferentes: un nivel muy estricto con $p=0.001$ y otro nivel con una $p=0.01$. El análisis de los datos empleando el análisis multivariante incluye también un cálculo de la tasa de posibles falsos positivos (False Discovery Rate o FDR). Los resultados de estos ensayos se muestran en la tabla 26. Aplicando el criterio más estricto, de $p=0.001$, se pueden identificar las diferentes diluciones de los péptidos Bmn como características de los sueros provenientes de paciente infectados por VIH. En el casos de los análisis empleando el criterio de $p=0.01$, se pueden identificar las diluciones de los péptidos Bmn, Bcon y F, manteniéndose los niveles de FDR en valores inferiores al 3.5% para los péptidos Bcon y F identificados.

Máx. 1 Falso Positivo con 99% de probabilidad			Test Multivariante con permutación p=0.001			Test Multivariante con permutación p=0.01		
Péptido	p-valor	FDR	Péptido	p-valor	FDR	Péptido	p-valor	FDR
Bmn 90	<1e-7	0.000	Bmn 90	<1e-7	0.000	Bmn 90	<1e-7	0.000
Bmn 45	<1e-7	0.000	Bmn 45	<1e-7	0.000	Bmn 45	<1e-7	0.000
Bmn 22.5	<1e-7	0.000	Bmn 22.5	<1e-7	0.000	Bmn 22.5	<1e-7	0.000
Bmn 11.25	3e-4	0.003	Bmn 11.25	3e-4	0.003	Bmn 11.25	3e-4	0.003
						BCon 90	0.007	0.032
						BCon 45	0.009	0.032
						BCon 22.5	0.007	0.032
						BCon 11.25	0.009	0.032
						F 45	0.007	0.032

Tabla 26 Resultados generados durante la caracterización de los péptidos más relevantes para discriminar los sueros positivos frente a los sueros negativos.

Al igual que en el caso de los análisis empleando los scatterplots, se han repetido las comparaciones entre los sueros de los diferentes subtipos frente a los sueros negativos identificándose en cada caso los péptidos que caracterizan de una forma estadísticamente significativa la respuesta frente a cada uno de los subtipos respecto a los sueros negativos. Los resultados de estas comparaciones empleando un análisis multivariante con permutación y $p=0.01$ están recogidos en la tabla 27.

El último tipo de comparaciones realizadas consistió en la identificación de perfiles de reactividad característicos para cada uno de los subtipos analizados. Para la realización de estas comparaciones al igual que en el caso anterior, se empleó un análisis multivariante con permutación y $p=0.01$. Los resultados de los ensayos de comparación en los que se enfrentaban cada uno de los subtipos frente al resto de los sueros positivos están recogidos en la tabla 28. Para los subtipos B, D, E y G es posible encontrar unos perfiles característicos que son diferentes de una forma estadísticamente significativa, mientras que en el caso de los subtipos, A, C, F no se puede identificar ningún perfil propio característico.

Subtipo	Péptido	p-valor	FDR	Diferencia en intensidad
E	A 11.25	2E-3	6E-3	2.19
	A 22.5	2E-3	6E-3	2.45
	A 45	2E-4	1E-3	3.85
	A 90	2E-4	1E-3	3.11
	C 11.25	2E-7	2E-6	8.80
	C 22.5	2E-3	6E-3	7.72
	C 45	1E-7	1E-7	9.88
	C 90	1E-7	1e.7	9.82
	F 11.25	9E-3	2E-2	2.44
	F 22.5	8E-3	2E-2	2.79
	F 45	1E-3	5E-3	5.03
	F 90	1E-3	5E-3	4.31
	E 11.25	3E-4	2E-3	3.43
	E 22.5	5E-3	1E-2	1.88
	E 45	1E-3	4E-3	2.34
C	Bmn 11.25	9E-4	2E-3	10.36
	Bmn 22.5	5E-5	1E-4	12.88
	Bmn 45	1E-5	1E-4	14.16
	Bmn 90	5E-5	2E-4	12.91
	BCon 11.25	1E-3	2E-4	2.66
	BCon 22.5	1E-4	8E-4	3.07
	BCon 45	4E-4	1E-3	2.50
	BCon 90	4E-4	1E-3	2.59
	C 11.25	2E-6	4E-5	3.77
	C 45	1E-6	4E-5	3.77
	C 90	2E-5	1E-4	3.89
	F 11.25	5E-3	9E-3	1.74
	F 22.5	3E-3	6E-3	1.80
	F 45	4E-4	2E-3	2.06
	F 90	1E-3	3E-3	2.12
	D 11.25	8E-5	4E-4	1.79
	D 22.5	8E-3	1E-2	1.40

Subtipo	Péptido	p-valor	FDR	Diferencia en intensidad
D	Bmn 11.25	5E-5	4E-4	21.42
	Bmn 22.5	1E-6	2E-5	23.73
	Bmn 45	6E-7	2E-5	24.18
	Bmn 90	3E-6	4E-5	23.34
	BCon 11.25	1E-3	4E-3	5.27
	BCon 22.5	5E-4	3E-3	5.07
	BCon 45	1E-3	4E-3	3.96
	BCon 90	6E-4	3E-3	3.83
	D 11.25	1E-3	4E-3	7.39
	D 22.5	3E-3	9E-3	4.95
	D 45	2E-3	6E-3	3.95
	D90	2E-3	7E-3	5.87
F	Bmn 11.25	2E-3	2E-2	8.77
	Bmn 22.5	1E-4	1E-3	12.42
	Bmn 45	7E-5	1E-3	15.67
	Bmn 90	1E-4	1E-3	14.91
	BCon 90	3E-3	2E-2	2.77
G	A 11.25	3E-05	2E-04	2.64
	A 22.5	9E-05	4E-04	2.92
	A 45	3E-04	8E-04	2.66
	A 90	1E-03	4E-03	2.70
	BCon 11.25	8E-03	1E-02	1.73
	BCon 22.5	2E-04	8E-04	2.04
	BCon 45	3E-05	2E-04	2.29
	BCon 90	3E-05	2E-04	2.69
	C 11.25	8E-05	4E-04	2.58
	C 45	1E-06	4E-05	2.46
	C 90	3E-06	5E-05	3.17
	D 11.25	2E-03	5E-03	1.50
	D 45	3E-03	6E-03	1.55
	E 22.5	1E-03	4E-03	0.48
	E 45	5E-03	8E-03	0.54

	D 45	5E-3	8E-3	1.42
A	Bmn 11.25	1E-3	7E-3	6.94
	Bmn 22.5	9E-5	1E-3	8.93
	Bmn 45	8E-5	1E-3	9.84
	Bmn 90	1E-4	2E-3	9.61
	BCon 11.25	2E-3	9E-3	1.69
	BCon 22.5	9E-4	7E-3	1.61
	BCon 45	3E-3	1E-2	1.44
	BCon 90	5E-3	1E-2	1.53
	C11.25	8E-3	2E-2	4.26
	C 90	5E-3	1E-2	3.96
	F 11.25	6E-3	1E-2	1.82
	F 22.5	6E-3	1E-2	2.26
	F 45	1E-3	9E-3	1.44

	F11.25	2E-03	4E-03	3.14
G	F 90	1E-03	4E-03	4.13
	Bmn 11.25	1E-03	4E-03	10.73
	Bmn 22.5	1E-04	5E-04	10.43
	Bmn 45	6E-05	3E-04	14.60
	Bmn 90	1E-04	5E-04	13.89
B	Bmn 11.25	2E-4	2E-3	12.68
	Bmn 22.5	1E-4	1E-3	13.53
	Bmn 45	1E-4	1E-3	13.55
	Bmn 90	1E-4	1E-3	12.52
	BCon 11.25	2E-3	1E-2	3.95
	BCon 22.5	2E-3	1E-2	3.99
	BCon 45	5E-3	2E-2	3.39
	BCon 90	5E-3	2E-2	3.41
	F 45	2E-3	1E-2	1.90

Tabla27 con los resultados de las comparaciones entre los diferentes subtipos y los sueros negativos. En esta tabla se presentan los péptidos que resultan estadísticamente significativos a la hora de diferenciar los diferentes subtipos de los sueros negativos.

Subtipo	Péptido	p-valor	FDR	Diferencia en intensidad
O	O90	0.001	<1e-7	12.19
	O45	0.001	<1e-7	10.53
A	No se pueden detectar péptidos significativos			
B	A 11.25	0.001	0.02	0.20
	A 22.5	0.002	0.02	0.18
	A 45	0.003	0.03	0.16
	A 90	0.005	0.04	0.20
C	No se pueden detectar péptidos significativos			
D	D 11.25	0.001	1e-5	2.52
	D 22.5	0.004	1e-5	2.090
	D 45	0.003	4e-5	2.30
	D 90	0.001	4e-5	2.44

Subtipo	Péptido	p-valor	FDR	Diferencia en intensidad
E	E 11.25	0.004	0.018	1.22
	E 45	0.002	0.008	0.93
	E 90	0.005	0.013	0.93
	C 11.25	0.004	0.008	1.12
	C 22.5	0.002	0.008	2.39
	C 45	0.002	0.008	2.50
	C 90	0.002	0.008	2.39
F	No se pueden detectar péptidos significativos			
G	E 22.5	0.001	0.079	0.18
	E 45	0.008	0.079	0.20
	E 90	0.012	0.084	0.25

Tabla28 con la caracterización de los péptidos estadísticamente significativos para la discriminación entre los sueros pertenecientes a diferentes subtipos de VIH

4.2.4 Comparación entre resultados obtenidos mediante técnicas ELISA y micromatrices

La técnica estándar de referencia para la caracterización de los sueros infectados por VIH son los ensayos ELISA. Por este motivo uno de los aspectos analizados fue la comparación de los resultados obtenidos mediante ensayos de tipo ELISA con los obtenidos mediante los ensayos sobre la superficie de la micromatriz. Se han analizado y comparado los resultados obtenidos mediante las dos técnicas para identificar las diferencias y las coincidencias en los perfiles de reactividad en las siguientes condiciones:

- Comparación de los perfiles de reactividad de cada suero obtenidos mediante técnicas de ELISA y micromatrices de material biológico.
- Comparación de los perfiles de reactividad frente a cada uno de los péptidos comparando los resultados de los ensayos ELISA y de micromatrices de todos los sueros.
- Comparación de los perfiles de reactividad entre sueros del mismo subtipo y similares características.

Las comparaciones se han analizado mediante la utilización de los índices de correlación obtenidos entre los perfiles generados para cada suero mediante las diferentes técnicas en diferentes situaciones y empleando como valor de referencia de las micromatrices para las comparaciones el valor de reactividad obtenido para la dilución 45 mg/l de cada uno de los péptidos analizados.

La primera de las comparaciones realizadas consistió en analizar los perfiles de reactividad de cada uno de los sueros mediante las dos técnicas empleadas. Los resultados de esta comparación están recogidos en la tabla 29. Esta comparación muestra que los sueros que presentan una mejor correlación en los valores obtenidos mediante ambas técnicas son los pertenecientes al subtipo B, mientras en el resto los valores de correlación que se obtienen son bajos o muy bajos mostrando que hay poca correlación en el comportamiento de los sueros entre ambas técnicas.

Descripción del Suero	Suero	Índice de Correlación ELISA/micromatriz
Subtipo B débiles en ELISA	elisaST-461	0.773
	elisaST-711	0.856
	elisaST-732	0.957
	elisaST-453	0.426
Subtipo B	elisaST-455	0.961
	elisaWW-301-17	0.880
	elisaST-661	0.708
	elisaWW-301-2	0.099
	elisaST-372	0.532

Negativos en ensayos ELISA	elisaST-440	0.594
	elisaST-670	-0.215
	elisaST-794	0.061
	elisaST-941	0.037
Subtipo A	elisaST-687	0.650
	elisaWW-301-19	0.698
	elisaWW-301-12	0.738
Subtipo D	elisaWW-301-23	0.278
	elisaWW-301-25	0.611
	elisaWW-301-26	0.569
Subtipo C	elisaWW-301-18	0.026
	elisaWW-301-28	0.553
	elisaWW-301-31	0.210
Subtipo G	elisaWW-301-3	-0.004
	elisaWW-301-8	0.558
Subtipo E (CRF01_A/E)	elisaWW-301-38	0.317
	elisaWW-301-39	0.794
	elisaWW-301-40	0.494
Subtipo F	elisaWW-301-47	0.709
	elisaWW-301-48	0.842
	elisaWW-301-49	0.957
Tipo O	elisaWW-301-1	0.699
Promedio		0.541
Mediana		0.602
Max		0.960
Min.		-0.210
DesvEst		0.327

Tabla 29 Índices de correlación de los perfiles de reactividad obtenidos mediante técnicas ELISA y micromatrices

El siguiente análisis que se realizó consistió en la comparación entre el perfil de respuesta generado entre todos los sueros frente a un determinado péptido usando micromatrices o ensayos ELISA. Los índices de correlación generados durante estos análisis están recogidos en la tabla 30.

Péptido	Índice de Correlación
A	0.64
Bcon	0.93
Bmn	0.50
C	0.97
D	0.90
E	0.81
F	0.77

Tabla30 Índices de correlación entre las respuestas obtenidas para cada uno de los péptidos analizados

Se puede apreciar en esta tabla que los resultados obtenidos entre ambas técnicas están correlacionados entre sí, a excepción de los resultados observados para los péptidos A y Bmn. En el caso de estos dos péptidos el índice de correlación que se observa no muestra que exista correlación entre los resultados obtenidos para ambas técnicas, siendo, en todo caso una correlación débil la existente para la respuesta comparada entre las dos técnicas para el

péptido A (Índice de correlación= 0.64) y no pudiéndose concluir nada respecto al resultado obtenido en el análisis del péptido Bmn (Índice de correlación = 0.50).

El último de los análisis realizados para la comparación entre los resultados generados en los ensayos ELISA y en los ensayos con micromatrices consistió en el análisis de los índices de correlación obtenidos entre sueros con características similares entre si. Las características analizadas fueron el subtipo del suero y la intensidad de la positividad (fuerte o débil) determinada a partir del ensayo patrón reconocido que es el ensayo ELISA. Los resultados de este tipo de comparación muestran una mayor correlación en los resultados obtenidos mediante la técnica basada en micromatrices. En la Tabla 31 se presentan los resultados de estas comparaciones agrupados por las comparaciones realizadas. La técnica estándar de ELISA solo presenta unos mejores resultados en el caso de los sueros B débiles analizados, en el resto de los casos los resultados observados son inferiores a los que se han generado en los ensayos realizados en el formato de micromatriz. Hay que destacar sobre estos resultados el caso de los sueros menos reactivos, dado que en este grupo de sueros no se observa ningún tipo de correlación en los ensayos ELISA mientras que en el caso de los ensayos sobre micromatriz si que es posible apreciar una correlación débil.

Sueros	Micromatriz	ELISA
B débiles	0.87	0.98
B fuertes	0.81	0.81
A	0.80	0.56
C	0.94	0.27
D	0.90	0.84
E	0.72	0.73
F	0.91	0.92
Negativos en ELISA	0.66	-0.13

Tabla 31 Índices de correlación entre los resultados de los distintos sueros analizados generados mediante micromatrices o ensayos ELISA

4.2.5 Análisis de perfiles de reactividad con micromatrices

En el diseño experimental que se ha realizado se han empleado diferentes diluciones de los péptidos analizados para ser cuantificados. Los análisis anteriores han considerado los resultados obtenidos para cada uno de los péptidos como observaciones independientes entre si y sin relación, pero gracias al diseño se pueden analizar los resultados considerando el comportamiento global de las diferentes diluciones mediante el análisis de los resultados considerando los parámetros generados en el ajuste de las rectas de reactividad frente a cada péptido de los sueros analizados. Considerando como parámetro de interés la pendiente de las rectas ajustadas, se han repetido los análisis realizados para la comparación de los sueros en función de su positividad y del subtipo para la identificación de los parámetros estadísticamente significativos que caracterizan estos sueros.

Por analogía con los análisis realizados previamente, la primera de las características estudiadas fue la diferencia entre los sueros positivos y negativos. La selección de las reactividades analizadas se realizó mediante ensayos multivariantes con permutación

empleando $p=0.01$ y 1000 permutaciones. Los resultados de este análisis comparativo permiten identificar la reactividad frente a los péptidos Bmn, F, C y Bcon como los más significativos a la hora de discriminar entre los sueros positivos y negativos. Además las tasas de posibles falsos positivos se mantienen en unos niveles inferiores a 0.001. Los resultados de este análisis están recogidos en la tabla 32. Estos resultados son similares a los obtenidos mediante el análisis de los resultados considerando cada una de las diluciones independientemente.

Péptido	p-valor	FDR	Variación (Neg./Pos.)
Bmn	6e-7	6e-6	0.12
Bcon	5e-4	1e-3	0.36
C	3e-4	1e-3	0.45
F	2e-4	1e-3	0.42

Tabla 32 Péptidos cuyas pendientes resultan estadísticamente significativas para la discriminación entre sueros positivos y negativos.

Una vez realizado el análisis comparativo entre los sueros positivos y negativos la siguiente etapa consistió en analizar las diferencias entre los sueros positivos pertenecientes a los diferentes subtipos. La técnica de análisis y los parámetros fueron los mismos que los empleados anteriormente, permitiendo identificar los péptidos cuya respuesta permite su mejor caracterización. Los resultados de estos análisis están recogidos en la tabla 33.

Subtipo	Péptido	p-valor	FDR	Diferencia en intensidad (Otros/clase analizada)
A	No se pueden detectar péptidos significativos			
B	Bcon	3e-3	0.03	0.33
C	E	6e-4	0.006	5.98
D	D	5e-7	5e-6	0.16
E	C	4.5e-4	4e-3	0.19
F	No se pueden detectar péptidos significativos			
G	No se pueden detectar péptidos significativos			

Tabla 33 con los resultados de los análisis para identificar aquellos péptidos cuyas pendientes resultan estadísticamente significativas para discriminar entre los diferentes subtipos.

Al igual que sucedió en el análisis en el que se consideraron las diferentes diluciones como entidades independientes, en este otro tipo de análisis, no resulta posible identificar una respuesta que permita discriminar de manera estadísticamente significativa entre los sueros de subtipo A y el resto de los sueros.

En el caso de los subtipos B, se da la única diferencia existente entre los análisis realizados anteriormente y el análisis basado en las pendientes. En el análisis comparativo entre los sueros de subtipo B y el resto de los sueros es posible identificar la respuesta frente al péptido Bcon como estadísticamente significativa, y con una mayor pendiente, tres veces superior, en los sueros de subtipo B que en el resto de los subtipos.

Para los sueros pertenecientes al subtipo C este tipo de análisis permite la identificación de la respuesta frente al péptido E como una respuesta significativa, a diferencia de lo que sucedía con el análisis individual de los péptidos en el que no fue posible encontrar péptidos cuya reactividad fuera estadísticamente diferente. La diferencia en la reactividad frente al

péptido E presenta una pendiente casi 6 veces menor frente a este péptido que el resto de los subtipos.

En el análisis de los sueros de subtipo D, al igual que sucediera con el análisis de los péptidos individuales es posible identificar la reactividad frente al péptido D como la respuesta significativa y diferencial entre este subtipo y el resto de subtipos, siendo la pendiente de la respuesta frente a este subtipo aproximadamente siete veces la pendiente del resto de los subtipos.

La comparación de los sueros de subtipo E (CRF01_A/E) frente al resto permite la identificación en este caso de la reactividad frente al péptido C como respuesta estadísticamente significativa, teniendo en este caso la reactividad frente a este péptido una pendiente aproximadamente 5 veces superior a la pendiente para este péptido que se detecta en el resto de los subtipos.

Al igual que en el caso de los análisis empleando los péptidos individuales, no es posible identificar para las muestras pertenecientes al subtipo F un perfil de reactividad característico. En el caso de las muestras pertenecientes al subtipo G empleando este tipo de análisis basado en los perfiles de reactividad de las 4 diluciones de los péptidos analizadas simultáneamente tampoco es posible identificar ninguna respuesta única y significativamente significativa.

5. Discusión

5.1. Análisis de las superficies para la fabricación de las micromatrices

Un aspecto fundamental para la utilización de micromatrices de material biológico es la selección de las estrategias de inmovilización del material biológico que se desea emplear durante los ensayos. Las propiedades de las superficies generadas durante los procesos de funcionalización tienen por lo tanto una importancia capital a la hora de determinar otros aspectos relacionados con la metodología experimental que se puede emplear y sobre los requerimientos que deben tener las sondas que se deseen inmovilizar sobre ellas. Atendiendo a sus características físico-químicas de las superficies es posible catalogar estas superficies como:

- Uniones no covalentes-electrostáticas. En este tipo de superficies el proceso de inmovilización se produce por la interacción electrostática entre cargas eléctricas positivas presentes sobre la superficie de la micromatriz y las cargas negativas del esqueleto fosfato de los ácidos nucleicos. Este tipo de uniones entre la superficie y las sondas inmovilizadas limita el uso de este tipo de superficies a ensayos de hibridación en los que la sonda no requiere de ninguna orientación particular. En este tipo de superficies las sondas no requieren de ningún tipo de modificación particular para permitir su inmovilización. Durante este trabajo se han empleado superficies generadas con poly-L-lisina y aminosilano como superficies que presentan cargas positivas para este tipo de inmovilización.
- Uniones covalentes no orientadas. El proceso de inmovilización en estas superficies genera enlaces químicos estables entre los grupos funcionales añadidos a la superficie en el proceso de funcionalización y los ácidos nucleicos que se desea inmovilizar. La principal característica de este tipo de superficies es que la inmovilización va a generar enlaces covalentes, por lo general a través de grupos fosfato o amino naturalmente presentes en el ácido nucleico sin requerir de su modificación, en cualquier posición en la que estos se encuentren. De este modo la sonda se inmoviliza sobre la superficie de forma que es imposible saber cual es la orientación o conformación de la misma. Las aplicaciones de este tipo de superficies, al igual que sucede con las uniones electrostáticas, están limitadas a ensayos de hibridación.
- Uniones covalentes orientadas. La utilización de estas superficies requiere la modificación de las sondas de ácidos nucleicos que se van a inmovilizar para que se genere un enlace covalente estable entre el ácido nucleico modificado, a través de su modificación, y los grupos reactivos presentes sobre la superficie de la micromatriz. Al ser la reacción de formación del enlace por la posición modificada, se conoce cual va a ser el punto de unión entre la sonda y la superficie, por lo que es posible orientar la sonda inmovilizada. Este tipo de modificación abre las puertas al desarrollo de diferentes aproximaciones basadas tanto en hibridación como en métodos enzimáticos en los que se requiere que una determinada región de las sondas esté accesible para

la enzima. Este tipo de aproximación suele requerir además, de la inclusión de una secuencia espaciadora (por lo general 10-15T) que aleje a la sonda de la superficie para disminuir posibles impedimentos estéricos con la superficie de la micromatriz. Las superficies de tipo epóxido e isotiocianato analizadas en este trabajo permiten este tipo de unión.

La comparación de superficies que se ha realizado en este trabajo ha estado dirigida a seleccionar las superficies que presentasen experimentalmente un mejor comportamiento para la realización tanto de ensayos de hibridación empleando sondas inmovilizadas no covalentemente como ensayos enzimáticos que requieren de la inmovilización covalente y orientada de las sondas. Este tipo de ensayo ha permitido la comparación entre superficies con inmovilización covalente orientada, inmovilización electrostática e inmovilización mixta, empleando una superficie que presentaba grupos para la inmovilización covalente y grupos cargados electrostáticamente (superficie amino-epoxisilano). Los criterios seleccionados para la determinación del mejor comportamiento fueron la reproducibilidad a la hora de fijar la misma cantidad de sonda en diferentes réplicas, la cantidad de sonda que son capaces de retener y el ruido de fondo de cada una de ellas mediante la realización de ensayos de hibridación bajo las mismas condiciones. Los resultados de estos ensayos han permitido identificar las superficies de tipo poly-L-lisina e isotiocianato como las dos superficies que presentan las mejores características. Estas dos superficies han sido las que en los ensayos se han comportado con una mayor homogeneidad, tanto en las intensidades observadas como en el tamaño de los puntos. También han sido estas dos superficies, junto con la superficie de tipo epóxido, las que han presentado un mejor comportamiento en términos de las intensidades de la señal y el ruido. En el caso de la superficie de tipo epóxido, se ha observado que es la que presenta una mayor capacidad de unión de sonda a la superficie lo cual la hacía presentarse como una buena candidata ya que puede usarse tanto empleando sondas modificadas como sondas sin modificar, pero en las condiciones que se han analizado este tipo de superficie generaba puntos más pequeños que el resto de las superficies presentando además una mayor dispersión en los valores de señal que se detectaron. Esta gran dispersión en los valores de señal observados ha sido el principal motivo para descartar la utilización de este tipo de superficie para los ensayos que se llevaron a cabo posteriormente.

Otro aspecto que se ha considerado en la realización de los ensayos de selección de las superficies ha sido el relacionado con el proceso de bloqueo de las mismas. El proceso de bloqueo consiste en el tratamiento de la superficie para inactivar los grupos funcionales que presentan en su superficie y que permiten la inmovilización del material biológico. Las observaciones realizadas han mostrado que en las condiciones de reacción empleadas no se observa diferencia entre las reacciones realizadas con micromatrices bloqueadas y sin bloquear. En el caso de la superficie de tipo isotiocianato, las micromatrices sufren un proceso de "auto bloqueo" al reaccionar los grupos isotiocianato con el CO₂ ambiental para generar una tiourea que impide posteriores uniones con las sondas.

5.2 Métodos de soporte al diseño de sondas y recuperación de secuencias

En la actualidad existe un gran número de herramientas bioinformáticas (Deng *et al.*, 2008; Feng and Tillier, 2007; Garbarine and Rosen, 2008; Li and Ying, 2006; Lin *et al.*, 2006; Rimour *et al.*, 2005; Wernersson and Nielsen, 2005; Gordon and Sensen, 2004) que permiten a los investigadores realizar el proceso de diseño de sondas para su inmovilización sobre la superficie de las micromatrices. La mayor parte de estas herramientas son para el diseño de sondas para ensayos de expresión génica, siendo menos frecuentes las herramientas dirigidas al diseño de sondas para otras aplicaciones. Como parte de este trabajo se ha realizado un sistema de soporte a la toma de decisiones para el diseño de sondas para experimentos con micromatrices. Este sistema ofrece al investigador un sistema que le guía a lo largo del proceso de diseño de sondas para diferentes aplicaciones de las micromatrices. Este proceso guiado, va detallando las diferentes etapas en las que se puede descomponer el proceso de diseño de sondas y ofrece enlaces a herramientas bioinformáticas (preferiblemente online, en los casos que ha sido posible) que permiten la realización de cada una de las etapas del proceso.

Un aspecto importante a la hora de diseñar sondas para desarrollar sistemas de detección de agentes microbiológicos empleando micromatrices es la correcta selección de la región sobre la que se desean localizar las sondas. En este tipo de aplicaciones es frecuente trabajar en las regiones definidas por las reacciones de PCR diagnósticas, debido a su elevada sensibilidad. En estos casos es necesario contar con una herramienta que permita identificar las regiones amplificadas en estas reacciones y recuperar las secuencias amplificadas para que sean analizadas o procesadas empleando otras herramientas posteriormente.

BUSSUB es una herramienta que ha sido diseñada con este propósito, y que a diferencia de otras herramientas (Rice *et al.*, 2000; Lexa *et al.*, 2001; Rotmistrovsky *et al.*, 2004) permite recuperar en grandes archivos de secuencias las subsecuencias definidas por los productos de PCR definidos por una pareja de cebadores que pueden incluir degeneraciones definidas según la codificación IUPAC y un número de posibles errores en las secuencias reconocidas definido por el usuario. Estas subsecuencias se recuperan en formato FASTA lo que permite su posterior utilización directa con otras herramientas bioinformáticas (por ejemplo para el alineamiento múltiple de secuencias o para el diseño de sondas para micromatrices). La definición por parte del usuario del número de errores reconocidos es diferente de la utilización de posiciones degeneradas. Este parámetro es especialmente útil durante el trabajo con secuencias de virus RNA en los que existe una elevada tasa de variabilidad que muchas veces no está bien cubierta por las posibles degeneraciones incluidas en los cebadores y que permite recuperar secuencias que no serían detectadas por otros métodos. En relación con estos posibles errores en las secuencias reconocidas respecto a los cebadores originales, se puede definir la severidad de las reacciones de PCR dado que el usuario puede definir una región a partir del extremo 3' del cebador en la cual no está permitido incluir ninguno de estos errores. Además de las características anteriormente citadas el programa permite también recuperar las secuencias "upstream" y "downstream" a partir de las

regiones identificadas como conservadas lo que puede ser de utilidad en estudios filogenéticos para analizar cual es la variabilidad en regiones contiguas a estas regiones conservadas.

5.3 Sistemas de detección de microorganismos patógenos mediante micromatrices

El desarrollo de las técnicas basadas en biología molecular ha permitido una mejora en las técnicas de detección y diagnóstico de las enfermedades infecciosas. La aplicación de sistemas basados en la utilización de micromatrices de material biológico permite aprovechar el potencial de muchas de las aplicaciones clínicas basadas en la utilización de reacciones de PCR de alta sensibilidad para combinarlas con la capacidad de detección sobre las micromatrices de un gran número de secuencias y características asociadas a las mismas.

En este trabajo se ha analizado la sensibilidad de este tipo de aproximación combinada y se han podido observar sensibilidades de alrededor de 10 copias de genoma completo del organismo que se desea analizar. El desarrollo de sistemas basados en este tipo de aproximaciones permite ahondar en la caracterización de los organismos que se están analizando, permitiendo en algunos casos la discriminación a nivel de especie o el genotipado de las mismas en base a las diferencias en las secuencias que son amplificadas. La puesta a punto de técnicas de amplificación multiplexadas en las que es posible amplificar varias regiones pertenecientes a un mismo organismo o a varios, se ve reforzada por la especificidad de la detección sobre las micromatrices de material biológico.

5.3.1 Detección de patógenos de interés en bioterrorismo

Actualmente es posible encontrar en los laboratorios de microbiología clínica ensayos basados en reacciones de biología molecular para la detección de agentes patógenos y susceptibles de ser empleados en acciones de bioterrorismo. El desarrollo de ensayos basados en micromatrices para la detección de este tipo de agentes patógenos, sirvió para la puesta a punto de las técnicas de fabricación, marcaje y reacción que se han empleado posteriormente también en los ensayos con muestras clínicas.

En el proceso de puesta a punto y desarrollo de la metodología necesaria para la realización de este tipo de ensayos, se ha podido observar que la inclusión de las moléculas marcadoras durante el proceso de marcaje tiene repercusiones en la eficiencia de las reacciones de amplificación. Al aumentar la concentración de análogo de nucleótido marcado se observa una reducción en la cantidad de producto de PCR generado. Esto es debido a que los grupos fluorescentes que están incluidos en los análogos de nucleótido presentan impedimentos estéricos en su interacción con la polimerasa lo que lleva a una reducción de la actividad de la misma. Este efecto sobre las reacciones de amplificación debe ser considerado a la hora de diseñar los ensayos puesto que puede provocar la modificación de las condiciones de PCR en aquellos casos en los que se desee pasar un ensayo basado en reacciones de PCR a uno en el que se desee hacer una detección posterior mediante técnicas basadas en micromatrices.

En los ensayos realizados para el diseño de una micromatriz capaz de detectar diversos agentes de interés en bioterrorismo, hay que destacar que fue necesaria la generación de una secuencia química artificial de un fragmento del virus de la viruela para poder realizar los ensayos de validación de los sistemas destinados a su detección. Esta secuencia química se generó mediante un proceso de reacciones de PCR anidadas y posterior clonación y secuenciación. Una selección cuidadosa de las condiciones de PCR y de las concentraciones de los oligonucleótidos empleados permitió simplificar este proceso y su realización en una única etapa.

Los ensayos de validación han demostrado que las micromatrices pueden constituir una herramienta para la detección específica de los agentes de interés empleando reacciones de hibridación. En el caso concreto de los ensayos realizados con muestras problema para la detección de *F. tularensis* en aguas contaminadas, los resultados observados se confirmaron por reacciones de PCR observándose únicamente un falso negativo en estos resultados. Este resultado se explica porque los análisis se realizaron en unas concentraciones que rozaron las condiciones de dilución límite.

En este trabajo se han puesto a punto protocolos que permiten aplicar estas técnicas para la detección de los agentes de interés a partir de productos de amplificación mediante la extensión de sondas diseñadas para tal efecto. Dentro de este entorno se ha explorado también el desarrollo de sistemas basados en la extensión de cebadores inmovilizados sobre la superficie de las micromatrices. Este tipo de abordaje presenta algunas ventajas e inconvenientes respecto a las técnicas basadas en hibridación. En comparación con las reacciones de hibridación, las reacciones de extensión de cebadores inmovilizados son más rápidas en términos de tiempo, el protocolo de extensión (30 minutos) representa un 25% comparado con el de hibridación (2 horas). Esta reducción de tiempo representa una ventaja en un entorno clínico en el que el factor tiempo es más importante que en los entornos de investigación. El desarrollo de sondas específicas para la realización de los ensayos de extensión de cebadores requiere analizar las características del problema dado que existen dos posibles aproximaciones, una de ellas en la que se emplean terminadores de cadena marcados fluorescentemente y otra en la que se emplean sondas alelo-específicas que son extendidas. En los ensayos realizados durante este trabajo, se ha mostrado que ambas aproximaciones son posibles. La selección de una u otra está determinada por el tipo de experimento que se desea realizar. La utilización de terminadores de cadena presenta como principal inconveniente una menor sensibilidad que la observada en los ensayos de extensión de sondas alelo específicas. Esto es debido a que en el caso de la utilización de sondas alelo específicas se incorpora más de una molécula fluorescente en las sondas inmovilizadas, mientras que en los ensayos en los que se emplean terminadores de cadena esta incorporación está limitada a una única molécula fluorescente por cada una de las sondas inmovilizadas. Este efecto tiene una importancia creciente a medida que se reduce la concentración de muestra que sirve de molde para la reacción de extensión. Hay que tener en cuenta que este tipo de ensayos son isotermos y que a diferencia de lo que sucede en las reacciones de PCR no se producen ciclos de

desnaturalización e hibridación entre la muestra y la sonda. En estos casos la diferencia en la cantidad de marcaje incorporada entre las dos aproximaciones va aumentando a medida que se reduce la concentración de muestra en el ensayo. Durante este trabajo se han intentado algunos ensayos dirigidos a la realización de reacciones de PCR sobre la superficie de la micromatriz o a la utilización de una estrategia con ciclos de desnaturalización/extensión para las reacciones de extensión, pero los resultados de los mismos fueron insatisfactorios (datos no mostrados). Se observó evaporación de las muestras y/o una degradación de la superficie de las micromatrices que provocaba un aumento del ruido de fondo en el ensayo que imposibilitaba la detección de cualquier tipo de señal.

5.3.2 Sistemas de detección de virus respiratorios

El desarrollo de un sistema basado en micromatrices para la detección y tipado de adenovirus respiratorios fue seleccionado como modelo para el desarrollo de una aplicación de diagnóstico clínico de la metodología desarrollada. Los nuevos métodos basados en la utilización de técnicas de la biología molecular se han ido convirtiendo en herramientas de diagnóstico rutinario para la detección y tipado de virus respiratorios. El desarrollo de métodos como la RT-PCR/PCR, la secuenciación, las micromatrices o las combinaciones entre ellos representan un avance, mejorando la velocidad y la sensibilidad respecto a otras técnicas clásicas o a los métodos basados en cultivos.

El conjunto de sondas específicas empleadas para la generación de la micromatriz desarrollada en este trabajo permite la detección e identificación de los principales serotipos de los adenovirus involucrados en infecciones respiratorias agudas (adenovirus de serotipos 1, 2, 3, 4, 5, 7, 14 y 21) y la detección genérica de cualquier adenovirus mediante una sonda especialmente diseñada para esta función. Las sondas se diseñaron para reconocer diferentes regiones de un fragmento amplificado del gen que codifica para la proteína Hexon. El dispositivo permite una identificación visual, sobre la imagen generada por el escáner, de los diferentes serotipos de interés detectados mediante el reconocimiento de diferentes patrones de hibridación únicos identificados para cada uno de ellos.

Los niveles de sensibilidad determinados para la detección positiva de un adenovirus presente en las muestras se fijaron en un rango de entre 10 y 100 moléculas de partida, mientras que para la identificación de los patrones característicos que permiten la caracterización y la asignación de un serotipo concreto se necesitan entre 100 y 1000 copias de partida. Los ensayos de especificidad de las reacciones sobre la superficie de las micromatrices se llevaron a cabo empleando otros virus respiratorios tales como influenza o RSV y un panel de referencia que constaba de los 49 serotipos de los adenovirus humanos, del cual se excluyeron los serotipos prevalentes en las infecciones agudas respiratorias. En los ensayos con los virus respiratorios diferentes de los adenovirus no se produjo en ningún caso hibridación. En el caso de los ensayos de especificidad empleando el panel de los serotipos de referencia, se detectó señal procedente de la sonda genérica para la detección de cualquier adenovirus en todas las hibridaciones, pudiendo observarse en algunas ocasiones algunas

señales muy débiles con alguna de las sondas específicas pero en ningún caso generando un patrón que pudiera confundirse con los generados por los serotipos con tropismo por el tracto respiratorio.

El proceso de validación clínica de los dispositivos y metodologías desarrolladas se realizó mediante el análisis de diez especímenes clínicos procedentes de aspirados nasofaríngeos de pacientes con infecciones respiratorias agudas. Estas muestras fueron inicialmente caracterizadas mediante otras técnicas como infecciones provocadas por adenovirus sin conocerse el serotipo del adenovirus causante de la infección. La asignación de las muestras analizadas a un serotipo concreto se realizó mediante la comparación de los patrones de hibridación observados en las muestras clínicas frente al patrón generado por los serotipos respiratorios de referencia empleados como control. Los resultados de estas identificaciones fueron posteriormente confirmados mediante la secuenciación y posterior análisis filogenético de las regiones amplificadas en estas muestras. En todos los casos los resultados obtenidos mediante los ensayos sobre las micromatrices se vieron respaldados por el análisis filogenético.

Los resultados de estos ensayos mostraron, como era de esperar, la detección e identificación de adenovirus con tropismo por el tracto respiratorio perteneciendo a los serotipos 2, 3, 4 y 5 que eran los serotipos prevalentes en el momento en el que se tomaron las muestras (Casas *et al.*, 2005).

El desarrollo de nuevas técnicas y herramientas diagnosticas era uno de los objetivos planteados en este trabajo, tratando de simplificar algunos de los procesos, técnicas y medios necesarios para la aplicación de las mismas, siendo necesario únicamente un termociclador, un horno de hibridación y un escáner para la lectura de las micromatrices. Además se ha simplificado al máximo el número de reactivos y procedimientos de análisis necesarios. Para facilitar la posible implantación de este tipo de técnicas en entornos clínicos se ha limitado el proceso a 4 etapas: Extracción de la muestra, Amplificación y marcaje, hibridación sobre la micromatriz (sin requerirse la purificación de las muestras o procesos de bloqueo o prehibridación de las micromatrices) y una última etapa de escaneado e interpretación visual de los datos.

5.4 Análisis de la respuesta inmune humoral frente diferentes subtipos de VIH

Las aplicaciones de las micromatrices de material biológico en entornos de microbiología clínica no están solo limitadas a la identificación de posibles agentes patógenos de interés mediante la detección de ácidos nucleicos (Burgess *et al.*, 2008; Ofir *et al.*, 2005). También pueden ser empleadas para el análisis de las infecciones desde otras perspectivas como puede ser la respuesta inmune frente a las mismas por parte de los sujetos infectados. Se han desarrollado varios trabajos dirigidos hacia la puesta a punto de sistemas análogos a las técnicas de ELISA para la detección de diferentes infecciones (Burgess *et al.*, 2008). Este tipo

de aproximaciones son por lo general similares a las desarrolladas en el entorno de los ácidos nucleicos para la detección de agentes infecciosos y están fundamentalmente basadas en aproximaciones cualitativas más que cuantitativas.

En este trabajo se ha seguido un enfoque alternativo basado en el análisis mediante técnicas basadas en micromatrices para analizar el problema del subtipado de VIH desde el punto de vista de la respuesta inmune de los pacientes frente a diferentes subtipos genéticos de este virus en contraposición a los estudios basados en el genotipado viral. En este tipo de abordaje se han inmovilizado sobre la superficie de las micromatrices péptidos sintéticos que caracterizan la región variable V3 de los diferentes subtipos de la proteína gp120 del VIH.

5.4.1 Desarrollo de estrategias de control de calidad y análisis de datos

A diferencia de las aplicaciones basadas en la utilización de micromatrices en las que se inmovilizan sondas de ácidos nucleicos que tienen cerca de dos décadas de existencia, los desarrollos para la inmovilización de proteínas son más recientes y están menos estandarizados. Por este motivo ha sido necesario desarrollar una estrategia completa para controlar la calidad del proceso de fabricación de las micromatrices y partiendo de unos datos reproducibles analizar y desarrollar estrategias de normalización que permitan la comparación de los datos.

Los experimentos se diseñaron de manera que sobre la superficie de las micromatrices se inmovilizaron cuatro diluciones seriadas de cada uno de los péptidos de interés. Esto permitió analizar la respuesta de los diferentes sueros y la linealidad de las respuestas medidas. Estos análisis mostraron la linealidad de las respuestas frente a los diferentes péptidos en los diferentes subtipos analizados. El desarrollo de este tipo de estrategias se puede emplear también como una herramienta más para el control de calidad de los experimentos permitiendo la eliminación de puntos anómalos o descartando enteramente resultados en los que existan valores frente a diversos péptidos que no sean lineales.

En el análisis de datos de expresión génica generados mediante técnicas de micromatrices se han desarrollado un gran número de métodos de normalización para la comparación de los experimentos. La mayoría de estas aproximaciones se basan en técnicas globales, en las que aplicando algoritmos de diferente complejidad se asume que los niveles globales de expresión van a ser similares y están relacionados mediante una constante de normalización. Una alternativa a este tipo de aproximación se basa en la inclusión de unos “spikes” en los experimentos que son unas muestras exógenas cuantificadas que se añaden en igual cantidad en todos los ensayos que se desea comparar de manera que la relación entre estos “spikes” se emplea como constante de normalización para la transformación de los resultados. En el caso de los experimentos realizados con las micromatrices de proteínas, cada muestra va a contener una diferente población de anticuerpos, tanto cualitativa como cuantitativamente, que van a reconocer de diferente manera los péptidos sintéticos inmovilizados sobre la superficie de la micromatriz. Por este motivo, en este tipo de ensayos no

se cumplen los parámetros en los que se fundamentan las técnicas globales de normalización. Por ello se desarrolló una técnica alternativa similar al concepto de inclusión de los “spikes”. Esta técnica consistió en la inmovilización sobre la superficie de la micromatriz de puntos que contienen dos concentraciones diferentes de IgG humana. Los anticuerpos anti-humanos marcados fluorescentemente que se encargan de reconocer las posiciones en las que ha habido reacciones positivas entre los sueros y los péptidos inmovilizados también reconocen estas IgG inmovilizadas de manera que los niveles de intensidad de las señales que se miden deberían ser iguales, por lo que es posible calcular una constante de normalización que permite la comparación entre los resultados de diferentes ensayos. Esta constante de normalización se genera para cada micromatriz haciendo que la mediana de los niveles de intensidad de los valores observados para la dilución “IgG 90” sean iguales a 1. Este tipo de transformación de los datos se mostró muy robusto y permitió el análisis comparativo cuantitativo de la respuesta inmune frente a diferentes subtipos.

5.4.2 Comparaciones genética vs. reconocimiento humoral

Uno de los objetivos de este trabajo era analizar la diferente respuesta humoral frente a la infección por diferentes subtipos del virus VIH-1, analizando la generación de anticuerpos capaces de reconocer diferentes péptidos sintéticos consenso pertenecientes a la región V3 de la proteína gp120 provenientes del análisis de numerosos genotipos asociados con los diferentes subtipos del HIV-1.

Los resultados de los ensayos realizados sobre las micromatrices permitieron la realización de diversos enfoques para el análisis de los datos en los que fue posible abordar la comparación numérica de las intensidades con las que los diferentes sueros reaccionaban frente a los péptidos inmovilizados sobre la superficie de las micromatrices.

En una primera aproximación para el análisis de los resultados se emplearon técnicas basadas en el análisis e interpretación gráfica de las diferencias localizadas entre los valores de intensidad observados para los sueros de pacientes seropositivos frente a HIV respecto a sueros de pacientes control seronegativos. Los resultados de estos análisis permitieron identificar principalmente la respuesta frente a los péptidos, Bmn, Bcon, C y F, como indicativa de seropositividad de las muestras. Para la realización de estos análisis, se calculó la media de las intensidades de los valores normalizados observados frente a cada una de las clases. La presencia de una mayoría de muestras procedentes de infecciones por subtipo B podría explicar los resultados y que estos tengan un gran parecido con los resultados provenientes del análisis de los sueros B aislados. La comparación de sueros de tipo O frente a los sueros negativos presentó, como era de esperar, el reconocimiento de los péptidos diseñados para el reconocimiento de este tipo de sueros. En el caso de la comparación de las reactividades de los sueros de tipo M frente a los de tipo O el resultado obtenido fue el esperado y similar a la superposición de las comparaciones entre las respuestas de ambos tipos comparadas con los sueros negativos. Se observó, que como era de esperar, los sueros de tipo O presentaban solo intensidades significativas en el reconocimiento de los péptidos asociados a este tipo de virus,

mientras que los sueros de tipo M presentaban unas intensidades significativas en el reconocimiento de los péptidos asociados al tipo M del virus. Al igual que sucedía en el caso de las comparaciones de sueros de pacientes infectados con virus de tipo M frente a los sueros negativos, solo se identificó la respuesta frente a los péptidos Bmn, Bcon, C y F como significativa. Inicialmente este resultado, al igual que en el caso de la comparación frente a los sueros negativos, era debido a la mayor presencia de sueros pertenecientes al subtipo B en los estudios. Posteriormente, la repetición de este tipo de análisis individualizado para cada uno de los subtipos frente a los sueros negativos mostró que, independiente del subtipo analizado, todos ellos presentaban una reactividad significativa frente a los péptidos Bmn y Bcon. La obtención de estos resultados resultó sorprendente dado que inicialmente se podría esperar que fuesen únicamente los péptidos asociados a cada uno de los subtipos analizados los que produjeran señal o al menos que esta señal fuese la mas intensa de las encontradas. Este tipo de comportamiento, en el cual se apreciaba una mayor reactividad de todos los sueros positivos frente a los péptidos Bmn ya se pudo apreciar al realizar los análisis de linealidad de las respuestas en los que sistemáticamente la respuesta frente a estos péptidos era superior a la respuesta frente al resto.

La segunda aproximación seguida para el análisis y la comparación de los resultados se basó en la utilización de técnicas estadísticas de comparación de clases para determinar, aquellos péptidos que permiten inicialmente la discriminación entre los sueros de tipo M de los sueros negativos, la discriminación de cada uno de los subtipos analizados frente a los sueros negativos y finalmente la discriminación entre sí de los diferentes subtipos. Los p-valores empleados para la realización de estos ensayos se mantuvieron en un nivel de 0.01 por las características de densidad de puntos de las micromatrices empleadas. Pese a tratarse de matrices con una baja densidad de puntos y a que solo se consideran 36 variables, se emplearon técnicas para la corrección y ajuste de los p-valores empleados en los ensayos de comparación. El resultado de este tipo de análisis empleando diferentes técnicas y p-valores para los análisis permitió, al igual que había sucedido inicialmente en los métodos de análisis gráficos, identificar la respuesta frente al péptido Bmn como la principal y más importante característica de los sueros de pacientes infectados por el virus del HIV. Al igual que se ha discutido en la interpretación gráfica de los resultados, inicialmente se atribuyó este mayor peso de los péptidos asociados al subtipo B como una consecuencia de su mayor abundancia entre los sueros analizados. Esta hipótesis se vio refrendada en el posterior análisis individualizado de los mismos cuando se pudo observar una vez más que la respuesta estadísticamente significativa frente al péptido Bmn aparecía en todos los subtipos analizados con la única excepción de los sueros pertenecientes al subtipo E (CRF01_A/E), siendo además ésta la única discrepancia observada entre los resultados analizados gráfica y estadísticamente.

El análisis comparativo entre los péptidos identificados de una forma estadísticamente significativa para la discriminación entre subtipos permitió identificar patrones de discriminación para 4/7 de los subtipos analizados, no siendo posible identificar péptidos que permitan

discriminar entre los sueros pertenecientes a los subtipos A, C y F. Para el resto de los subtipos, B, D, E (CRF01_A/E) y G si es posible identificar respuestas variables que permiten su identificación. En el caso de los sueros pertenecientes al subtipo B su principal característica es una caída en los niveles de intensidad que se observan en el reconocimiento de las diluciones del péptido asociado al subtipo A respecto al resto de los subtipos analizados. Para este subtipo inicialmente se habría esperado que las intensidades características fuesen debidas a un elevado reconocimiento de los péptidos Bmn o Bcon, que son los péptidos inicialmente caracterizados como propios del subtipo B. Como se ha observado en las comparaciones entre sueros positivos para la infección por HIV-1 respecto a sueros negativos, la respuesta a estos péptidos se aprecia en todos los subtipos, a excepción de los pertenecientes al subtipo E (CRF01_A/E). Por este motivo el sistema identificó como característico de los sueros de subtipo B esta reducción a un 20% en la reactividad frente al péptido A respecto al resto de los subtipos. En el caso de los sueros de subtipo E (CRF01_A/E), se ha identificado como característica la respuesta frente a los péptidos C y E. En este caso se puede apreciar un incremento significativo en la respuesta frente estos péptidos respecto a la respuesta del resto de los sueros, siendo el resultado inicialmente esperado el reconocimiento del péptido asociado al subtipo E (CRF01_A/E). Para el reconocimiento de los sueros de subtipo D se puede apreciar que es la respuesta frente a este péptido, tal y como sería de esperar, la que se ve incrementada y caracteriza a este subtipo. En el caso de los sueros pertenecientes al subtipo G se ha observado que su principal característica es una disminución a un 20% de la reactividad frente al péptido E en el conjunto formado por el resto de los subtipos. En este caso, hay que ser especialmente cuidadosos a la hora de interpretar los resultados para este subtipo dado que solo se pudieron analizar 2 sueros.

En el caso de los subtipos A, C y F no se pudieron identificar perfiles característicos. En el caso de los subtipos A y C, se realizaron diferentes comparaciones de sus respuestas, eliminando del conjunto de datos a comparar los sueros de los subtipos A o C según sea el subtipo comparado y E (CRF01_A/E), dado que filogenéticamente se considera que estos subtipos son los más próximos entre si, no habiendo sido posible la identificación en ninguna de estas condiciones de un perfil propio para estos subtipos. En el caso de los sueros de subtipo F, los péptidos que reconoce cuando se realiza la comparación frente a los sueros negativos son los péptidos B, que son unos péptidos reconocidos por el resto de los subtipos lo que provoca que no sea posible caracterizar una respuesta individual frente a este subtipo.

La existencia de diferencias entre los niveles genotípicos y fenotípicos ya se había observado anteriormente (Murphy *et al.*, 1999) y el análisis y la interpretación de los resultados obtenidos ha mostrado que no existe una correlación clara entre el genotipo viral, su fenotipo y por lo tanto la respuesta inmune asociada en la región analizada. Estas observaciones abren las puertas a la aplicación de métodos diferentes basados en técnicas para el descubrimiento de clases como pudieran ser las técnicas de “clustering” para agrupar los sueros en función de los valores de reactividad observados y analizados.

5.4.3 Comparación de resultados de ELISA vs. Micromatrices

Los resultados obtenidos en los ensayos empleando las micromatrices fueron comparados con los resultados obtenidos para los mismos sueros empleando técnicas de ELISA. Pese a que en ambas técnicas los péptidos están inmovilizados sobre una superficie, las densidades de material inmovilizado y las propiedades físico-químicas de los ensayos no tienen porque permanecer constantes al reducirse la escala del ensayo. Por lo tanto se llevaron a cabo comparaciones entre el comportamiento de los ensayos en una escala macroscópica como son los ensayos de tipo ELISA y los ensayos en escala microscópica sobre las micromatrices en los que se inmovilizaron los mismos péptidos.

Las comparaciones realizadas para analizar las similitudes y divergencias entre estas dos técnicas se fundamentaron en el análisis del comportamiento de los sueros individualmente frente a todos los péptidos y en el de las respuestas obtenidas frente a cada uno de los péptidos por el conjunto de los sueros en ambas técnicas. Este abordaje permitió realizar la comparación entre ambas técnicas desde los dos ejes que se emplearon en el estudio, sueros y péptidos. En los ensayos ELISA analizados se emplea un único pocillo para cada péptido, por lo que para la comparación se seleccionó la dilución 45 de los péptidos para llevar a cabo las comparaciones.

La comparación en el reconocimiento de los diferentes sueros mostró una gran variabilidad dependiendo de diferentes parámetros. Se pudo observar que los sueros que presentan una mejor correlación entre los resultados de micromatrices y ELISA son los sueros pertenecientes a los subtipos B y F mientras que en el resto de los sueros analizados hay poca o ninguna correlación entre los sueros. En el caso de los sueros B estos se subdividieron a su vez entre sueros con una fuerte señal en ELISA y sueros débiles en ELISA, los resultados muestran mayor correlación entre los sueros B débiles en ELISA que en los fuertemente positivos, entre estos cabe destacar la excepción del suero ww-301-2 cuyos ensayos no muestran ninguna correlación. A la vista de estos resultados, se repitió el ensayo ELISA y se observó que el suero había sido mal clasificado tratándose en realidad de un suero débil en ELISA. Pese a todo los resultados continuaron sin presentar correlación con los observados en micromatrices. Los sueros que presentan unos mejores resultados de correlación son aquellos que se corresponden con los subtipos en los que se esperaría observar unos valores más altos de señal en el reconocimiento de los péptidos B y F mientras que en el resto de los sueros pertenecientes al resto de subtipos (A, C, D, E y G) no se aprecia apenas correlación en los resultados. Como se ha indicado anteriormente se observó que generalmente para casi todos los sueros y subtipos las señales más fuertes se observaban en la reactividad frente a los péptidos Bmn y Bcon en los ensayos con micromatrices mientras que esto no fue así en el caso de los ensayos en formato ELISA.

En el caso de los ensayos en los que se analizó la correlación en la respuesta frente a los sueros individualmente se pudo observar que existe una correlación ente los valores

analizados para los péptidos Bcon, C, D, E y F mientras que la correlación cae fuertemente en la comparación de la respuesta entre los péptidos A y Bmn.

Una posible explicación para estas discrepancias se basa en la imposibilidad de “normalizar” o transformar los datos de los diferentes ensayos ELISA para su comparación dado que cada determinación se realiza en un pocillo diferente y en unas condiciones macroscópicamente similares pero sin la certeza de que a nivel microscópico las condiciones realmente sean similares o equivalentes para poder tratar los datos como si fuesen realmente comparables entre sí. El desarrollo de estrategias de normalización y transformación de los datos en los estudios con micromatrices ha sido y sigue siendo un área en el que se publican multitud de trabajos, lo que da una idea de su importancia. Al no existir ni poderse desarrollar tan siquiera una normalización básica de los datos de los ensayos ELISA se deben considerar con cuidado aquellos datos en los que no se observa correlación o esta es muy baja, dado que es imposible discriminar si esta falta de correlación es debida a diferencias entre las dos técnicas o si por el contrario las diferencias son artefactos producidos por la utilización de unos datos no normalizados procedentes de los ensayos ELISA.

Otro aspecto que se ha tenido también en cuenta a la hora de analizar los datos ha sido la correlación de los datos obtenidos para sueros que presentaban características comunes como el genotipo o el título de anticuerpos presente en estos sueros. Este tipo de comparaciones muestra que los ensayos realizados sobre las micromatrices presentan correlación entre sí cuando se analizan en base a estas características mientras que la repetición de los ensayos analizando los datos procedentes de los ensayos ELISA presenta en un 25% de las condiciones analizadas una correlación muy baja o ausencia de correlación entre las muestras descritas como pertenecientes al mismo grupo.

5.4.4 Caracterización de perfiles antigénicos mediante ensayos de “afinidad”

Para la realización de los ensayos se diseñaron micromatrices en las que cada péptido estaba representado por cuatro concentraciones diferentes. Como se indicó anteriormente, las cuatro concentraciones están comprendidas dentro de un rango lineal de respuesta, este comportamiento nos ha permitido desarrollar un nuevo tipo de abordaje para el análisis de los datos no descrito hasta ahora en el que se puede incluir el concepto de “afinidad” de un suero frente a los péptidos inmovilizados sobre la superficie de las micromatrices. Esta “afinidad” se podría definir a partir de la pendiente de la recta ajustada que se genera al analizar la respuesta de un suero frente a las cuatro diluciones de un péptido simultáneamente. De esta manera se podría decir que cuando se comparan dos sueros reactivos frente a un mismo péptido tiene una mayor afinidad por ese péptido el que presenta una pendiente mayor al ajustarse las rectas. Una vez definida esta “afinidad” se puede emplear, al igual que sucedió anteriormente con las diluciones individuales de los péptidos, para comparar y caracterizar los sueros pertenecientes a los diferentes subtipos.

El análisis de las comparaciones realizadas siguiendo este método arroja unos resultados similares a los observados durante el análisis con las diluciones de los péptidos. En el caso de las comparaciones entre los sueros negativos y positivos se puede apreciar que en el caso de los sueros HIV positivos estos presentan una mayor afinidad por los péptidos Bmn, Bcon, C y F al igual que se observó en los anteriores análisis. Las diferencias surgen al caracterizar las posibles diferencias entre los subtipos. En este caso, empleando el análisis de “afinidad” encontramos que como inicialmente sería de esperar estos sueros presentan una mayor “afinidad” por el péptido Bcon independientemente de la intensidad con la que se produce el reconocimiento. En el análisis empleando independientemente las cuatro diluciones de los péptidos se observó que la principal característica de este suero era un menor reconocimiento, traducido en una menor intensidad de señal frente a los péptidos A, observándose por lo tanto una discrepancia entre los resultados obtenidos por ambas técnicas de análisis de datos. Estas discrepancias se observan también en el caso de los sueros pertenecientes al subtipo C, en el cual es posible identificar mediante este método una característica propia que es la menor afinidad por el péptido E respecto al resto de los subtipos, y para el subtipo G en este caso no se puede apreciar ningún tipo de respuesta característica. En el caso de los subtipos A y F la utilización de un método alternativo para la caracterización de los sueros sigue sin arrojar ninguna luz acerca de características propias para cada uno de estos subtipos a nivel de la respuesta inmune que generan en las condiciones analizadas. En los sueros pertenecientes a los subtipos E se observa una situación de coincidencia intermedia con los resultados generados por el otro método de análisis de los datos, siendo en este caso lo más significativo el importante aumento de la “afinidad” por el péptido C lo que le distingue del resto, mientras que se pierde el reconocimiento del péptido E como característica discriminante. En los sueros de subtipo D tal y como se observó en los análisis previos la característica definitoria de este subtipo es el reconocimiento del péptido D.

5.4.5 Interpretación de los resultados de afinidad frente a los resultados empleando los péptidos independientemente.

La diferencia en la interpretación de los resultados entre los ensayos empleando métodos en los que se analizan los datos basados en “afinidad” o métodos en los que se analiza la intensidad de las señales independientemente está en que en un caso, cuando se emplean los cuatro péptidos individualmente se está analizando la intensidad global con la que un suero reconoce un determinado péptido y está relacionada con la abundancia de moléculas capaces de interactuar con ese péptido. En el caso de los ensayos en los que se analiza la “afinidad” la intensidad del reconocimiento de cada uno de los péptidos no es relevante, si no que se considera la forma en la que aumenta esa intensidad al aumentar la cantidad de antígeno disponible, estando este concepto por lo tanto más próximo al de la especificidad de la reacción de reconocimiento entre las especies moleculares involucradas en el análisis.

6. Conclusiones

Conclusiones

A partir de los resultados obtenidos en este trabajo se puede concluir que:

1. Para los objetivos planteados en este trabajo las superficies basadas en isotiocianato y poli-l-lisina ofrecen las mejores prestaciones para la inmovilización de las sondas.
2. La incorporación de análogos de nucleótido marcados fluorescentemente reduce la eficiencia de las reacciones de PCR.
3. Las micromatrices de material biológico se pueden usar para la detección de agentes patógenos susceptibles de ser empleados en bioterrorismo (específicamente las clases A y B de la clasificación del CDC) con una sensibilidad de hasta 10 copias de genoma.
4. La aplicación de herramientas bioinformáticas y micromatrices de material biológico posibilita el serotipado de adenovirus respiratorios humanos en muestras clínicas.
5. La herramienta bioinformática BUSSUB permite la realización de reacciones de PCR “in Silico” y facilita el diseño de sondas para su utilización en micromatrices.
6. La respuesta inmune humoral frente a diferentes subtipos del VIH-1 puede ser caracterizada empleando micromatrices y técnicas bioinformáticas, poniendo de manifiesto que existen diferencias entre la caracterización genética y la caracterización serológica.
7. Los péptidos Bmn y Bcon permiten la discriminación entre sueros de pacientes infectados por el VIH-1 de tipo M frente a sueros de pacientes no infectados.
8. Los resultados obtenidos mediante técnicas de ELISA y técnicas basadas en micromatrices no presentan correlación en la respuesta frente a los péptidos individuales analizados.
9. Los resultados obtenidos con micromatrices presentan una mejor correlación entre los valores de la respuesta inmune asociados con diferentes características fenotípicas y genotípicas comunes entre los sueros estudiados que los ensayos ELISA.
10. La utilización de micromatrices de material biológico permite estudiar la “afinidad”, entendida como la pendiente de la respuesta medida frente a un péptido, de los sueros de pacientes infectados por diferentes subtipos del VIH-1 para su caracterización.

7. Resumen Final

En los últimos años el desarrollo de nuevas tecnologías y nuevas técnicas en la biología molecular han abierto las puertas al desarrollo de nuevas aplicaciones en campos tales como la salud pública y el diagnóstico microbiológico. Así mismo la bioinformática se ha consolidado como un campo y una disciplina fundamental para el almacenamiento e interpretación de los resultados de los ensayos de laboratorio y la gestión del conocimiento generado en las ciencias biomédicas. En este entorno se ha desarrollado el trabajo recogido en esta tesis doctoral en la cual se han empleado técnicas de biología molecular y bioinformática para la generación de un conjunto innovador de métodos y herramientas dirigidas hacia entornos de aplicación en microbiología clínica pero que pueden ser exportadas hacia otras posibles aplicaciones. Las técnicas desarrolladas en este trabajo se sitúan en tres grandes áreas:

1. Generación de herramientas bioinformáticas dirigidas a facilitar el proceso de diseño de sondas para su utilización en micromatrices de ácidos nucleicos y para el análisis "*in Silico*" de reacciones de PCR
2. Desarrollo de un entorno basado en micromatrices de material biológico para la detección de agentes infecciosos causantes de patologías humanas y análisis de muestras clínicas empleando los dispositivos generados.
3. Estudio y caracterización de la respuesta inmune humoral frente a la infección por el VIH-1 mediante micromatrices de proteínas.

Para el proceso de desarrollo de dispositivos basados en la utilización de las micromatrices para la detección de agentes microbiológicos ha sido necesaria la investigación y desarrollo de soluciones para numerosos problemas relacionados con estas técnicas. Ha sido necesario analizar y estudiar aspectos relacionados con la generación de los sustratos sobre los que se produce la inmovilización de las moléculas biológicas que se van a emplear como sondas para reconocer los agentes de interés. El estudio comparativo entre las diferentes superficies analizadas ha permitido la identificación de las superficies de tipo poly-L-lisina e isotiocianato como las superficies más recomendables de entre las analizadas para ser empleadas en el desarrollo de estos dispositivos. La utilización de estas superficies se adecua muy bien al entorno de aplicación al que van dirigidas ya que ambas son capaces de inmovilizar una gran diversidad de material biológico con unas características complementarias entre si y que permiten el desarrollo de numerosas aplicaciones posteriores basadas tanto en ensayos enzimáticos sobre la superficie de las micromatrices como de afinidad (por ejemplo hibridaciones de ácidos nucleicos o reacciones antígeno anticuerpo). Una vez fue seleccionado el soporte de la micromatriz el siguiente paso dado para el desarrollo de los dispositivos consistió en el diseño y validación de las sondas capaces de reconocer, para su identificación, a los agentes de interés. El proceso de diseño de sondas está relacionado con el tipo de ensayo a realizar. Para la realización de este trabajo se han desarrollado ensayos basados en el reconocimiento mediante hibridación de ácidos nucleicos y ensayos de extensión de cebadores inmovilizados sobre la superficie de las micromatrices. En el proceso de desarrollo de los sistemas de detección de micromatrices se ha incluido el desarrollo de sistemas

bioinformáticos de apoyo al proceso de diseño de sondas. Se ha desarrollado una herramienta online de ayuda a la toma de decisiones para el proceso de diseño de sondas y se ha desarrollado una herramienta, BUSSUB, para facilitar la selección de las regiones que deben ser empleadas durante el proceso de diseño de sondas. BUSSUB también puede ser empleado para el análisis de reacciones de PCR *"in Silico"*, identificando los posibles amplicones generados por un par de cebadores.

Inicialmente se seleccionó la detección de agentes susceptibles de ser empleados en bioterrorismo y de interés en biodefensa como el entorno sobre el cual desarrollar las técnicas y metodologías necesarias. Se diseñó y generó un dispositivo con sondas inmovilizadas sobre su superficie para la detección simultánea de varios organismos de interés (*Bacillus anthracis*, *Burkholderia mallei*, *Burkholderia pseudomallei*, *Burkholderia thailandensis* y otros miembros del género *Burkholderia*, *Brucella abortus*, *Brucella melitensis*, *Brucella canis*, *Brucella suis* y otros miembros del género *Brucella*, *Coxiella burnetii* y otros miembros del género *Coxiella*, *Francisella philomiragia*, *Francisella tularensis* *tularensis* y otros miembros del género *Francisella*, virus *Variola* y *Yersinia pestis*). La inclusión del virus *variola* motivó la necesidad de diseñar una quimera sintética que contuviese una secuencia del virus de la viruela para ser empleada durante los ensayos de validación de las sondas diseñadas y la utilización de muestras problema generadas en el laboratorio para la validación de las mismas y de las técnicas desarrolladas. Los ensayos de validación de este dispositivo mostraron una elevada sensibilidad y especificidad siendo posible la detección de hasta 50 copias de genoma de los organismos de interés. La traslación de los desarrollos generados en este modelo a un entorno de aplicación diagnóstica empleando muestras clínicas reales se hizo mediante la aplicación de técnicas basadas en ensayos de hibridación sobre micromatrices para la detección y caracterización de adenovirus respiratorios (adenovirus de tipo 1, 2, 3, 4, 5, 7, 14 y 21). Para el desarrollo de los dispositivos y de las metodologías empleadas en estos ensayos se aplicaron parte de las técnicas generadas en el entorno modelo de detección de agentes de interés en biodefensa y bioterrorismo y se desarrollaron nuevas sondas específicas para la detección de los virus de interés. El desarrollo de este dispositivo de aplicación clínica permitió la identificación y genotipado en diez muestras clínicas de los adenovirus respiratorios presentes en las mismas mediante el reconocimiento de patrones de hibridación característicos para cada uno de los adenovirus de interés. Los resultados de la caracterización de las muestras sobre las micromatrices fue refrendada mediante la posterior secuenciación y análisis filogenético de las mismas.

El análisis de los datos generados por la aplicación de un sistema de micromatrices para el análisis de la respuesta inmune humoral frente a la infección por VIH-1, ha proporcionado un escenario integrado para la aplicación de todas estas técnicas y metodologías. En el caso de los desarrollos realizados para la generación de dispositivos y herramientas dirigidas a la detección de microorganismos a nivel de ácidos nucleicos los análisis realizados han sido de tipo cualitativo, dirigidos a una determinación positiva o negativa de la presencia de un determinado agente en una muestra. El análisis de la respuesta inmune ha sido un análisis de

tipo cuantitativo ya que no solo se ha analizado si un determinado suero presenta reactividad frente a unos péptidos sintéticos, generados a partir de la región V3 de la proteína gp120 del VIH-1 y característicos de diversos subtipos de este virus, si no que se ha analizado el grado de reactividad de ese suero frente a todos y cada uno de esos péptidos. Este tipo de análisis ha permitido identificar el reconocimiento del péptido Bmn como característica de una infección por VIH-1 independientemente del subtipo viral que haya causado la infección primaria. También ha sido posible en algunos casos identificar unos perfiles de reactividad de los sueros asociados a los diferentes subtipos, en los que se ha observado que no hay una correlación entre los resultados del subtipado genético y los resultados del subtipado inmunológico, dado que por lo general no es la respuesta frente al péptido característico de ese subtipo la responsable de caracterizar y permitir la discriminación entre los diferentes subtipos.

El análisis comparativo entre los resultados obtenidos mediante micromatrices y técnicas de ELISA mostró que, si bien la respuesta a nivel de los diferentes sueros presentaba una baja correlación entre los resultados de ambas técnicas, el análisis individualizado de los diferentes péptidos mostró que se podía encontrar correlación en los resultados de ambas técnicas en aproximadamente un 75% de los péptidos analizados. Además se pudo observar también que los resultados agrupando los sueros estudiados en función de diferentes características fenotípicas y genotípicas presentaban una mayor correlación en los ensayos realizados en formato de micromatriz.

Gracias al diseño de los ensayos ha sido posible proponer una nueva visión y enfoque para el análisis de este tipo de datos. La utilización de diluciones seriadas ha permitido analizar la respuesta desde un punto de vista de la “afinidad” de un determinado suero frente a los diferentes péptidos. La “afinidad” se ha definido como la pendiente de la recta generada al analizar de una forma conjunta las intensidades observadas del reconocimiento de cada uno de los péptidos. El desarrollo de este nuevo tipo de análisis proporciona un enfoque diferente a la interpretación de los resultados generados en este tipo de ensayos, dado que los análisis “clásicos” están enfocados a interpretar los datos en función de los niveles de intensidad de la señal (cantidad de anticuerpos que interaccionan en cada una de las posiciones analizadas) mientras que el nuevo enfoque basado en la “afinidad” se basa en un concepto diferente, analizando cómo varía la señal al aumentar la cantidad de antígeno sobre la muestra. Este nuevo concepto permite una interpretación diferente de los datos ya que analiza los datos desde otra perspectiva.

La aplicación conjunta de las nuevas tecnologías basadas en micromatrices de material biológico y bioinformática se ha mostrado como una estrategia adecuada para el desarrollo de nuevas aplicaciones para la identificación y caracterización de agentes infecciosos y para el estudio de las interacciones entre estos patógenos y el sistema inmune humano.

7 Bibliografía

Bibliografía

- Abzug, M.J., Beam, A.C., Gyorkos, E.A., and Levin, M.J. (1990). Viral pneumonia in the first month of life. *Pediatr. Infect. Dis. J.* 9, 881-885.
- Abzug, M.J. and Levin, M.J. (1991). Neonatal adenovirus infection: four patients and review of the literature. *Pediatrics* 87, 890-896.
- Allard, A., Albinsson, B., and Wadell, G. (2001). Rapid typing of human adenoviruses by a general PCR combined with restriction endonuclease analysis. *J. Clin. Microbiol.* 39, 498-505.
- Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990). Basic local alignment search tool. *J. Mol. Biol.* 215, 403-410.
- Avellon, A., Perez, P., Aguilar, J.C., Lejarazu, R., and Echevarria, J.E. (2001). Rapid and sensitive diagnosis of human adenovirus infections by a generic polymerase chain reaction. *J. Virol. Methods* 92, 113-120.
- Bao, Y., Bolotov, P., Dernovoy, D., Kiryutin, B., Zaslavsky, L., Tatusova, T., Ostell, J., and Lipman, D. (2008). The influenza virus resource at the National Center for Biotechnology Information. *J. Virol.* 82, 596-601.
- Barre-Sinoussi, F., Chermann, J.C., Rey, F., Nugeyre, M.T., Chamaret, S., Gruest, J., Dauguet, C., xler-Blin, C., Vezinet-Brun, F., Rouzioux, C., Rozenbaum, W., and Montagnier, L. (1983). Isolation of a T-lymphotropic retrovirus from a patient at risk for acquired immune deficiency syndrome (AIDS). *Science* 220, 868-871.
- Barrett, T., Troup, D.B., Wilhite, S.E., Ledoux, P., Rudnev, D., Evangelista, C., Kim, I.F., Soboleva, A., Tomashevsky, M., Marshall, K.A., Phillippy, K.H., Sherman, P.M., Muerter, R.N., and Edgar, R. (2009). NCBI GEO: archive for high-throughput functional genomic data. *Nucleic Acids Res.* 37, D885-D890.
- Beare, P.A., Unsworth, N., Andoh, M., Voth, D.E., Omsland, A., Gilk, S.D., Williams, K.P., Sobral, B.W., Kupko, J.J., III, Porcella, S.F., Samuel, J.E., and Heinzen, R.A. (2009). Comparative genomics reveal extensive transposon-mediated genomic plasticity and diversity among potential effector proteins within the genus *Coxiella*. *Infect. Immun.* 77, 642-656.
- Belsy, A., Odalys, V., Alexander, P., Clara, S., Angel, G., Grehete, G., Guelsys, G., Luis, S., Pedro, M., Guadalupe, G.M., Alina, L., Pilar, P.B., and Inmaculada, C. (2009). Molecular characterization of adenoviral infections in Cuba: report of an unusual association of species D adenoviruses with different clinical syndromes. *Arch. Virol.* 154, 619-627.
- Bracho, M.A., Gosalbes, M.J., Blasco, D., Moya, A., and Gonzalez-Candelas, F. (2005). Molecular epidemiology of a hepatitis C virus outbreak in a hemodialysis unit. *J. Clin. Microbiol.* 43, 2750-2755.
- Bracho, M.A., Gosalbes, M.J., Gonzalez, F., Moya, A., and Gonzalez-Candelas, F. (2006). Molecular epidemiology and evolution in an outbreak of fulminant hepatitis B virus. *J. Clin. Microbiol.* 44, 1288-1294.
- Brazma, A., Hingamp, P., Quackenbush, J., Sherlock, G., Spellman, P., Stoeckert, C., Aach, J., Ansorge, W., Ball, C.A., Causton, H.C., Gaasterland, T., Glenisson, P., Holstege, F.C., Kim, I.F., Markowitz, V., Matese, J.C., Parkinson, H., Robinson, A., Sarkans, U., Schulze-Kremer, S., Stewart, J., Taylor, R., Vilo, J., and Vingron, M. (2001). Minimum information about a microarray experiment (MIAME)-toward standards for microarray data. *Nat. Genet.* 29, 365-371.
- Buitrago Serna, M.J., Casas, F., I, Eiros-Bouza, J.M., Escudero, N.R., Giovanni, F.C., Jado, G., I, Pozo, S.F., Rubio Munoz, J.M., Sanchez-Seco Farinas, M.P., Valdezate, R.S., and Verdejo, O.J.

- (2007). [Biodefense: a new challenge for microbiology and public health]. *Enferm. Infecc. Microbiol. Clin.* 25, 190-198.
- Burgess, S.T., Kenyon, F., O'Looney, N., Ross, A.J., Kwan, M.C., Beattie, J.S., Petrik, J., Ghazal, P., and Campbell, C.J. (2008). A multiplexed protein microarray for the simultaneous serodiagnosis of human immunodeficiency virus/hepatitis C virus infection and typing of whole blood. *Anal. Biochem.* 382, 9-15.
- Carballal, G., Videla, C., Misirlian, A., Requeijo, P.V., and Aguilar, M.C. (2002). Adenovirus type 7 associated with severe and fatal acute lower respiratory infections in Argentine children. *BMC. Pediatr.* 2, 6.
- Casas, I., Avellon, A., Mosquera, M., Jabado, O., Echevarria, J.E., Campos, R.H., Rewers, M., Perez-Brena, P., Lipkin, W.I., and Palacios, G. (2005). Molecular identification of adenoviruses in clinical samples by analyzing a partial hexon genomic region. *J. Clin. Microbiol.* 43, 6176-6182.
- Chen, G.Y., Uttamchandani, M., Zhu, Q., Wang, G., and Yao, S.Q. (2003). Developing a strategy for activity-based detection of enzymes in a protein microarray. *Chembiochem.* 4, 336-339.
- Coiras, M.T., Perez-Brena, P., Garcia, M.L., and Casas, I. (2003). Simultaneous detection of influenza A, B, and C viruses, respiratory syncytial virus, and adenoviruses in clinical samples by multiplex reverse transcription nested-PCR assay. *J. Med. Virol.* 69, 132-144.
- Coker, R. (2009). Swine flu. *BMJ* 338, b1791.
- de Lang A., Baas, T., Teal, T., Leijten, L.M., Rain, B., Osterhaus, A.D., Haagmans, B.L., and Katze, M.G. (2007). Functional genomics highlights differential induction of antiviral pathways in the lungs of SARS-CoV-infected macaques. *PLoS. Pathog.* 3, e112.
- Deng, Y., He, Z., Van Nostrand, J.D., and Zhou, J. (2008). Design and analysis of mismatch probes for long oligonucleotide microarrays. *BMC. Genomics* 9, 491.
- Derbes, V.J. (1966). De Mussis and the great plague of 1348. A forgotten episode of bacteriological warfare. *JAMA* 196, 59-62.
- DeRisi, J., Penland, L., Brown, P.O., Bittner, M.L., Meltzer, P.S., Ray, M., Chen, Y., Su, Y.A., and Trent, J.M. (1996). Use of a cDNA microarray to analyse gene expression patterns in human cancer. *Nat. Genet.* 14, 457-460.
- Dewan, P.K., Fry, A.M., Laserson, K., Tierney, B.C., Quinn, C.P., Hayslett, J.A., Broyles, L.N., Shane, A., Winthrop, K.L., Walks, I., Siegel, L., Hales, T., Semenova, V.A., Romero-Steiner, S., Elie, C., Khabbaz, R., Khan, A.S., Hajjeh, R.A., and Schuchat, A. (2002). Inhalational anthrax outbreak among postal workers, Washington, D.C., 2001. *Emerg. Infect. Dis.* 8, 1066-1072.
- Dittmar, M.T., Simmons, G., Hibbitts, S., O'Hare, M., Louisirirotchanakul, S., Beddows, S., Weber, J., Clapham, P.R., and Weiss, R.A. (1997). Langerhans cell tropism of human immunodeficiency virus type 1 subtype A through F isolates derived from different transmission groups. *J. Virol.* 71, 8008-8013.
- Domingo, C., Palacios, G., Niedrig, M., Cabrerizo, M., Jabado, O., Reyes, N., Lipkin, W.I., and Tenorio, A. (2004). A New Tool for the Diagnosis and Molecular Surveillance of Dengue Infections in Clinical Samples. *Dengue Bulletin* 28, 87-95.
- Drosten, C., Gunther, S., Preiser, W., van der, W.S., Brodt, H.R., Becker, S., Rabenau, H., Panning, M., Kolesnikova, L., Fouchier, R.A., Berger, A., Burguiere, A.M., Cinatl, J., Eickmann, M., Escriou, N., Grywna, K., Kramme, S., Manuguerra, J.C., Muller, S., Rickerts, V., Sturmer, M., Vieth, S., Klenk, H.D., Osterhaus, A.D., Schmitz, H., and Doerr, H.W. (2003). Identification of a novel coronavirus in patients with severe acute respiratory syndrome. *N. Engl. J. Med.* 348, 1967-1976.

- Ebner,K., Rauch,M., Preuner,S., and Lion,T. (2006). Typing of human adenoviruses in specimens from immunosuppressed patients by PCR-fragment length analysis and real-time quantitative PCR. *J. Clin. Microbiol.* 44, 2808-2815.
- Echavarria,M. (2008). Adenoviruses in immunocompromised hosts. *Clin. Microbiol. Rev.* 21, 704-715.
- Echavarria,M., Forman,M., Ticehurst,J., Dumler,J.S., and Charache,P. (1998a). PCR method for detection of adenovirus in urine of healthy and human immunodeficiency virus-infected individuals. *J. Clin. Microbiol.* 36, 3323-3326.
- Echavarria,M., Forman,M., Ticehurst,J., Dumler,J.S., and Charache,P. (1998b). PCR method for detection of adenovirus in urine of healthy and human immunodeficiency virus-infected individuals. *J. Clin. Microbiol.* 36, 3323-3326.
- Echavarria,M., Forman,M., van Tol,M.J., Vossen,J.M., Charache,P., and Kroes,A.C. (2001). Prediction of severe disseminated adenovirus infection by serum PCR. *Lancet* 358, 384-385.
- Elrifro,E.M., Cooper,R.J., Dady,I., Hany,S., Mughal,Z.M., and Klapper,P.E. (2005). Three nonfatal cases of neonatal adenovirus infection. *J. Clin. Microbiol.* 43, 5814-5815.
- Engelthaler,D.M., Gage,K.L., Monteneri,J.A., Chu,M., and Carter,L.G. (1999). PCR detection of *Yersinia pestis* in fleas: comparison with mouse inoculation. *J. Clin. Microbiol.* 37, 1980-1984.
- Esteban,J.I., Gomez,J., Martell,M., Cabot,B., Quer,J., Camps,J., Gonzalez,A., Otero,T., Moya,A., Esteban,R., and . (1996). Transmission of hepatitis C virus by a cardiac surgeon. *N. Engl. J. Med.* 334, 555-560.
- Faix,D.J., Houg,h.S., Gaydos,J.C., Liu,S.K., Connors,J.T., Brown,X., Asher,L.V., Vaughn,D.W., and Binn,L.N. (2004). Evaluation of a rapid quantitative diagnostic test for adenovirus type 4. *Clin. Infect. Dis.* 38, 391-397.
- Falsey,J.R., Renil,M., Park,S., Li,S., and Lam,K.S. (2001). Peptide and small molecule microarray for high throughput cell adhesion and functional assays. *Bioconjug. Chem.* 12, 346-353.
- Feng,S. and Tillier,E.R. (2007). A fast and flexible approach to oligonucleotide probe design for genomes and gene families. *Bioinformatics.* 23, 1195-1202.
- Filliol,I., Motiwala,A.S., Cavatore,M., Qi,W., Hazbon,M.H., Bobadilla,d., V, Fyfe,J., Garcia-Garcia,L., Rastogi,N., Sola,C., Zozio,T., Guerrero,M.I., Leon,C.I., Crabtree,J., Angiuoli,S., Eisenach,K.D., Durmaz,R., Joloba,M.L., Rendon,A., Sifuentes-Osornio,J., Ponce de,L.A., Cave,M.D., Fleischmann,R., Whittam,T.S., and Alland,D. (2006). Global phylogeny of *Mycobacterium tuberculosis* based on single nucleotide polymorphism (SNP) analysis: insights into tuberculosis evolution, phylogenetic accuracy of other DNA fingerprinting systems, and recommendations for a minimal standard SNP set. *J. Bacteriol.* 188, 759-772.
- Fleischmann,R.D., Adams,M.D., White,O., Clayton,R.A., Kirkness,E.F., Kerlavage,A.R., Bult,C.J., Tomb,J.F., Dougherty,B.A., Merrick,J.M., and . (1995). Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. *Science* 269, 496-512.
- Flcury,H.J., Toni,T., Lan,N.T., Hung,P.V., Deshpande,A., Recordon-Pinson,P., Boucher,S., Lazaro,E., Jauvin,V., Lavignolle-Aurillac,V., Lebel-Binay,S., Cheret,A., and Masquelier,B. (2006). Susceptibility to antiretroviral drugs of CRF01_AE, CRF02_AG, and subtype C viruses from untreated patients of Africa and Asia: comparative genotypic and phenotypic data. *AIDS Res. Hum. Retroviruses* 22, 357-366.
- Fodor,S.P., Read,J.L., Pirrung,M.C., Stryer,L., Lu,A.T., and Solas,D. (1991). Light-directed, spatially addressable parallel chemical synthesis. *Science* 251, 767-773.

Forsman,M., Sandstrom,G., and Sjostedt,A. (1994). Analysis of 16S ribosomal DNA sequences of Francisella strains and utilization for determination of the phylogeny of the genus and for identification of strains by PCR. *Int. J. Syst. Bacteriol.* **44**, 38-46.

Freiberg,C. and Brotz-Oesterhelt,H. (2005). Functional genomics in antibacterial drug discovery. *Drug Discov. Today* **10**, 927-935.

Garbarine,E. and Rosen,G. (2008). An information theoretic method of microarray probe design for genome classification. *Conf. Proc. IEEE Eng Med. Biol. Soc.* **2008**, 3779-3782.

Garcia Del,B.N., Dobson,M.E., Vela,A.I., De,L.P., V, Gutierrez,C.B., Hadfield,T.L., Kuhnert,P., Frey,J., Dominguez,L., and Rodriguez Ferri,E.F. (2002). Genotyping of Francisella tularensis strains by pulsed-field gel electrophoresis, amplified fragment length polymorphism fingerprinting, and 16S rRNA gene sequencing. *J. Clin. Microbiol.* **40**, 2964-2972.

Garcia Pelayo,M.C., Uplekar,S., Keniry,A., Mendoza,L.P., Garnier,T., Nunez,G.J., Boschiroli,L., Zhou,X., Parkhill,J., Smith,N., Hewinson,R.G., Cole,S.T., and Gordon,S.V. (2009). A comprehensive survey of single nucleotide polymorphisms (SNPs) across Mycobacterium bovis strains and M. bovis BCG vaccine strains refines the genealogy and defines a minimal set of SNPs that separate virulent M. bovis strains and M. bovis BCG strains. *Infect. Immun.* **77**, 2230-2238.

Garcia-Albert, L., Lopez-Campos, G. H., Martin-Sanchez, F., and Garcia-Saiz, A. Desarrollo de un sistema para detección de anticuerpos frente al dominio V3 del VIH-1 mediante microarrays de péptidos sintéticos. Correlación con resultados de ELISA. VIII Congreso Nacional de Virología . 12-10-2003.

Ref Type: Conference Proceeding

Garcia-Albert,L., Ortiz,M., and Garcia-Saiz,A. (2001). HIV type 1 non-B subtype prevalence in Spain, 1997-1998. *AIDS Res. Hum. Retroviruses* **17**, 1317-1320.

Gardner,P.S., McGuckin,R., and McQuillin,J. (1972). Adenovirus demonstrated by immunofluorescence. *Br. Med. J.* **3**, 175.

Gordon,P.M. and Sensen,C.W. (2004). Osprey: a comprehensive tool employing novel methods for the design of oligonucleotides for DNA sequencing and microarrays. *Nucleic Acids Res.* **32**, e133.

Gray,G.C. (2006). Adenovirus transmission--worthy of our attention. *J. Infect. Dis.* **194**, 871-873.

Gray,G.C., Setterquist,S.F., Jirsa,S.J., DesJardin,L.E., and Erdman,D.D. (2005). Emergent strain of human adenovirus endemic in Iowa. *Emerg. Infect. Dis.* **11**, 127-128.

Gueudin,M., Plantier,J.C., Lemee,V., Schmitt,M.P., Chartier,L., Bourlet,T., Ruffault,A., Damond,F., Vray,M., and Simon,F. (2007). Evaluation of the Roche Cobas TaqMan and Abbott RealTime extraction-quantification systems for HIV-1 subtypes. *J. Acquir. Immune. Defic. Syndr.* **44**, 500-505.

Guo,Z., Guilfoyle,R.A., Thiel,A.J., Wang,R., and Smith,L.M. (1994). Direct fluorescence analysis of genetic polymorphisms by hybridization with oligonucleotide arrays on glass supports. *Nucleic Acids Res.* **22**, 5456-5465.

Haab,B.B., Dunham,M.J., and Brown,P.O. (2001). Protein microarrays for highly parallel detection and quantitation of specific proteins and antibodies in complex solutions. *Genome Biol.* **2**, RESEARCH0004.

Haase,A., Melder,A., Smith-Vaughan,H., Kemp,D., and Currie,B. (1995). RAPD analysis of isolates of Burkholderia pseudomallei from patients with recurrent melioidosis. *Epidemiol. Infect.* **115**, 115-121.

- Hao,K., Schadt,E.E., and Storey,J.D. (2008). Calibrating the performance of SNP arrays for whole-genome association studies. *PLoS. Genet.* 4, e1000109.
- Hartmann,M., Roeraade,J., Stoll,D., Templin,M.F., and Joos,T.O. (2009a). Protein microarrays for diagnostic assays. *Anal. Bioanal. Chem.* 393, 1407-1416.
- Hartmann,M., Roeraade,J., Stoll,D., Templin,M.F., and Joos,T.O. (2009b). Protein microarrays for diagnostic assays. *Anal. Bioanal. Chem.* 393, 1407-1416.
- Heim,A., Ebnet,C., Harste,G., and Pring-Akerblom,P. (2003). Rapid and quantitative detection of human adenovirus DNA by real-time PCR. *J. Med. Virol.* 70, 228-239.
- Hierholzer,J.C. (1992). Adenoviruses in the immunocompromised host. *Clin. Microbiol. Rev.* 5, 262-274.
- Holguin,A., Alvarez,A., and Soriano,V. (2005). Heterogeneous nature of HIV-1 recombinants spreading in Spain. *J. Med. Virol.* 75, 374-380.
- Holguin,A., Alvarez,A., and Soriano,V. (2002). HIV-1 subtype J recombinant viruses in Spain. *AIDS Res. Hum. Retroviruses* 18, 523-529.
- Holguin,A., Rodes,B., and Soriano,V. (2000). Recombinant human immunodeficiency viruses type 1 circulating in Spain. *AIDS Res. Hum. Retroviruses* 16, 505-511.
- Holmes,E.C. and Burch,S.S. (2000). The causes and consequences of genetic variation in dengue virus. *Trends Microbiol.* 8, 74-77.
- Holmes,K.V. (2003). SARS-associated coronavirus. *N. Engl. J. Med.* 348, 1948-1951.
- Hotopp,J.C., Grifantini,R., Kumar,N., Tzeng,Y.L., Fouts,D., Frigimelica,E., Draghi,M., Giuliani,M.M., Rappuoli,R., Stephens,D.S., Grandi,G., and Tettelin,H. (2006). Comparative genomics of *Neisseria meningitidis*: core genome, islands of horizontal transfer and pathogen-specific genes. *Microbiology* 152, 3733-3749.
- Hubble,J., Demeter,J., Jin,H., Mao,M., Nitzberg,M., Reddy,T.B., Wymore,F., Zachariah,Z.K., Sherlock,G., and Ball,C.A. (2009). Implementation of GenePattern within the Stanford Microarray Database. *Nucleic Acids Res.* 37, D898-D901.
- Julg,B. and Goebel,F.D. (2005). HIV genetic diversity: any implications for drug resistance? *Infection* 33, 299-301.
- Klinger,J.R., Sanchez,M.P., Curtin,L.A., Durkin,M., and Matyas,B. (1998). Multiple cases of life-threatening adenovirus pneumonia in a mental health care center. *Am. J. Respir. Crit Care Med.* 157, 645-649.
- Knudsen,T.B. and Daston,G.P. (2005). MIAME guidelines. *Reprod. Toxicol.* 19, 263.
- Kondo,M., Shima,T., Nishizawa,M., Sudo,K., Iwamuro,S., Okabe,T., Takebe,Y., and Imai,M. (2005). Identification of attenuated variants of HIV-1 circulating recombinant form 01_AE that are associated with slow disease progression due to gross genetic alterations in the nef/long terminal repeat sequences. *J. Infect. Dis.* 192, 56-61.
- Kumar,S., Tamura,K., and Nei,M. (2004). MEGA3: Integrated software for Molecular Evolutionary Genetics Analysis and sequence alignment. *Brief. Bioinform.* 5, 150-163.
- Kurg,A., Tonisson,N., Georgiou,I., Shumaker,J., Tollett,J., and Metspalu,A. (2000). Arrayed primer extension: solid-phase four-color DNA resequencing and mutation detection technology. *Genet. Test.* 4, 1-7.
- Laibson,P.R. (1975). Adenoviral keratoconjunctivitis. *Int. Ophthalmol. Clin.* 15, 187-201.

Lander, E.S., Linton, L.M., Birren, B., Nusbaum, C., Zody, M.C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., Funke, R., Gage, D., Harris, K., Heaford, A., Howland, J., Kann, L., Lehoczký, J., LeVine, R., McEwan, P., McKernan, K., Meldrim, J., Mesirov, J.P., Miranda, C., Morris, W., Naylor, J., Raymond, C., Rosetti, M., Santos, R., Sheridan, A., Sougnez, C., Stange-Thomann, N., Stojanovic, N., Subramanian, A., Wyman, D., Rogers, J., Sulston, J., Ainscough, R., Beck, S., Bentley, D., Burton, J., Clee, C., Carter, N., Coulson, A., Deadman, R., Deloukas, P., Dunham, A., Dunham, I., Durbin, R., French, L., Grafham, D., Gregory, S., Hubbard, T., Humphray, S., Hunt, A., Jones, M., Lloyd, C., McMurray, A., Matthews, L., Mercer, S., Milne, S., Mullikin, J.C., Mungall, A., Plumb, R., Ross, M., Shownkeen, R., Sims, S., Waterston, R.H., Wilson, R.K., Hillier, L.W., McPherson, J.D., Marra, M.A., Mardis, E.R., Fulton, L.A., Chinwalla, A.T., Pepin, K.H., Gish, W.R., Chissoe, S.L., Wendl, M.C., Delehaunty, K.D., Miner, T.L., Delehaunty, A., Kramer, J.B., Cook, L.L., Fulton, R.S., Johnson, D.L., Minx, P.J., Clifton, S.W., Hawkins, T., Branscomb, E., Predki, P., Richardson, P., Wenning, S., Slezak, T., Doggett, N., Cheng, J.F., Olsen, A., Lucas, S., Elkin, C., Uberbacher, E., Frazier, M., Gibbs, R.A., Muzny, D.M., Scherer, S.E., Bouck, J.B., Sodergren, E.J., Worley, K.C., Rives, C.M., Gorrell, J.H., Metzker, M.L., Naylor, S.L., Kucherlapati, R.S., Nelson, D.L., Weinstock, G.M., Sakaki, Y., Fujiyama, A., Hattori, M., Yada, T., Toyoda, A., Itoh, T., Kawagoe, C., Watanabe, H., Totoki, Y., Taylor, T., Weissenbach, J., Heilig, R., Saurin, W., Artiguenave, F., Brottier, P., Bruls, T., Pelletier, E., Robert, C., Wincker, P., Smith, D.R., Doucette-Stamm, L., Rubenfield, M., Weinstock, K., Lee, H.M., Dubois, J., Rosenthal, A., Platzer, M., Nyakatura, G., Taudien, S., Rump, A., Yang, H., Yu, J., Wang, J., Huang, G., Gu, J., Hood, L., Rowen, L., Madan, A., Qin, S., Davis, R.W., Federspiel, N.A., Abola, A.P., Proctor, M.J., Myers, R.M., Schmutz, J., Dickson, M., Grimwood, J., Cox, D.R., Olson, M.V., Kaul, R., Raymond, C., Shimizu, N., Kawasaki, K., Minoshima, S., Evans, G.A., Athanasiou, M., Schultz, R., Roe, B.A., Chen, F., Pan, H., Ramser, J., Lehrach, H., Reinhardt, R., McCombie, W.R., de la, B.M., Dedhia, N., Blocker, H., Hornischer, K., Nordsiek, G., Agarwala, R., Aravind, L., Bailey, J.A., Bateman, A., Batzoglou, S., Birney, E., Bork, P., Brown, D.G., Burge, C.B., Cerutti, L., Chen, H.C., Church, D., Clamp, M., Copley, R.R., Doerks, T., Eddy, S.R., Eichler, E.E., Furey, T.S., Galagan, J., Gilbert, J.G., Harmon, C., Hayashizaki, Y., Haussler, D., Hermjakob, H., Hokamp, K., Jang, W., Johnson, L.S., Jones, T.A., Kasif, S., Kasprzyk, A., Kennedy, S., Kent, W.J., Kitts, P., Koonin, E.V., Korf, I., Kulp, D., Lancet, D., Lowe, T.M., McLysaght, A., Mikkelsen, T., Moran, J.V., Mulder, N., Pollara, V.J., Ponting, C.P., Schuler, G., Schultz, J., Slater, G., Smit, A.F., Stupka, E., Szustakowski, J., Thierry-Mieg, D., Thierry-Mieg, J., Wagner, L., Wallis, J., Wheeler, R., Williams, A., Wolf, Y.I., Wolfe, K.H., Yang, S.P., Yeh, R.F., Collins, F., Guyer, M.S., Peterson, J., Felsenfeld, A., Wetterstrand, K.A., Patrinos, A., Morgan, M.J., de, J.P., Catanese, J.J., Osoegawa, K., Shizuya, H., Choi, S., and Chen, Y.J. (2001). Initial sequencing and analysis of the human genome. *Nature* 409, 860-921.

Larranaga, C., Kajon, A., Villagra, E., and Avendano, L.F. (2000). Adenovirus surveillance on children hospitalized for acute lower respiratory infections in Chile (1988-1996). *J. Med. Virol.* 60, 342-346.

Lexa, M., Horak, J., and Brzobohaty, B. (2001). Virtual PCR. *Bioinformatics.* 17, 192-193.

Li, W. and Ying, X. (2006). Mprobe 2.0: computer-aided probe design for oligonucleotide microarray. *Appl. Bioinformatics.* 5, 181-186.

Lin, B., Vora, G.J., Thach, D., Walter, E., Metzgar, D., Tibbetts, C., and Stenger, D.A. (2004). Use of oligonucleotide microarrays for rapid detection and serotyping of acute respiratory disease-associated adenoviruses. *J. Clin. Microbiol.* 42, 3232-3239.

Lin, F.M., Huang, H.D., Chang, Y.C., Tsou, A.P., Chan, P.L., Wu, L.C., Tsai, M.F., and Horng, J.T. (2006). Database to dynamically aid probe design for virus identification. *IEEE Trans. Inf. Technol. Biomed.* 10, 705-713.

Lindroos, K., Liljedahl, U., Raitio, M., and Syvanen, A.C. (2001). Minisequencing on oligonucleotide microarrays: comparison of immobilisation chemistries. *Nucleic Acids Res.* 29, E69.

- Long, G.W., Oprandy, J.J., Narayanan, R.B., Fortier, A.H., Porter, K.R., and Nacy, C.A. (1993). Detection of *Francisella tularensis* in blood by polymerase chain reaction. *J. Clin. Microbiol.* **31**, 152-154.
- Lopez-Campos, G., Coiras, M., Sanchez-Merino, J.P., Lopez-Huertas, M.R., Spiteri, I., Martin-Sanchez, F., and Perez-Brena, P. (2007). Oligonucleotide microarray design for detection and serotyping of human respiratory adenoviruses by using a virtual amplicon retrieval software. *J. Virol. Methods* **145**, 127-136.
- Lopez-Campos, G., Lopez, A., V, and Martin-Sanchez, F. (2008). Addressing the biomedical informatics needs of a microarray laboratory in a clinical microbiology context. *Stud. Health Technol. Inform.* **136**, 45-50.
- Lopez-Campos, G.H., Garcia-Albert, L., Martin-Sanchez, F., and Garcia-Saiz, A. (2006). Analysis and management of HIV peptide microarray experiments. *Methods Inf. Med.* **45**, 158-162.
- MacBeath, G. (2002). Protein microarrays and proteomics. *Nat. Genet.* **32 Suppl**, 526-532.
- MacBeath, G. and Schreiber, S.L. (2000). Printing proteins as microarrays for high-throughput function determination. *Science* **289**, 1760-1763.
- Macula, A.J., Schliep, A., Bishop, M.A., and Renz, T.E. (2008). New, improved, and practical k-stem sequence similarity measures for probe design. *J. Comput. Biol.* **15**, 525-534.
- Malasig, M.D., Goswami, P.R., Crawford-Miksza, L.K., Schnurr, D.P., and Gray, G.C. (2001). Simplified microneutralization test for serotyping adenovirus isolates. *J. Clin. Microbiol.* **39**, 2984-2986.
- Markowitz, V.M., Ivanova, N.N., Szeto, E., Palaniappan, K., Chu, K., Dalevi, D., Chen, I.M., Grechkin, Y., Dubchak, I., Anderson, I., Lykidis, A., Mavromatis, K., Hugenholtz, P., and Kyrpides, N.C. (2008). IMG/M: a data management and analysis system for metagenomes. *Nucleic Acids Res.* **36**, D534-D538.
- Maskos, U. and Southern, E.M. (1993). A study of oligonucleotide reassociation using large arrays of oligonucleotides synthesised on a glass support. *Nucleic Acids Res.* **21**, 4663-4669.
- Maskos, U. and Southern, E.M. (1992). Oligonucleotide hybridizations on glass supports: a novel linker for oligonucleotide synthesis and hybridization properties of oligonucleotides synthesised in situ. *Nucleic Acids Res.* **20**, 1679-1684.
- Maurer, M., Molitor, R., Sturn, A., Hartler, J., Hackl, H., Stocker, G., Prokesch, A., Scheideler, M., and Trajanoski, Z. (2005). MARS: microarray analysis, retrieval, and storage system. *BMC Bioinformatics.* **6**, 101.
- Mazurier, S., van de, G.A., Heuvelman, K., and Wernars, K. (1992). RAPD analysis of *Campylobacter* isolates: DNA fingerprinting without the need to purify DNA. *Lett. Appl. Microbiol.* **14**, 260-262.
- Mehra, S. and Kaushal, D. (2009). Functional Genomics Reveals Extended Roles of the *Mycobacterium tuberculosis* stress response factor σ^H . *J. Bacteriol.*
- Morales Espinosa, M.R., Delgado, S.G., and Cravioto, A. (2003). The use of microarrays for studying the pathogenesis of *Helicobacter pylori*. *Rev. Latinoam. Microbiol.* **45**, 24-29.
- Morfin, F., Dupuis-Girod, S., Frobert, E., Mundweiler, S., Carrington, D., Sedlacek, P., Bierings, M., Cetkovsky, P., Kroes, A.C., van Tol, M.J., and Thouvenot, D. (2009). Differential susceptibility of adenovirus clinical isolates to cidofovir and ribavirin is not related to species alone. *Antivir. Ther.* **14**, 55-61.

- Morfin,F., Dupuis-Girod,S., Mundweiler,S., Falcon,D., Carrington,D., Sedlacek,P., Bierings,M., Cetkovsky,P., Kroes,A.C., van Tol,M.J., and Thouvenot,D. (2005). In vitro susceptibility of adenovirus to antiviral drugs is species-dependent. *Antivir. Ther.* 10, 225-229.
- Muller,R., Ditzen,A., Hille,K., Stichling,M., Ehricht,R., Illmer,T., Ehninger,G., and Rohayem,J. (2009). Detection of herpesvirus and adenovirus co-infections with diagnostic DNA-microarrays. *J. Virol. Methods* 155, 161-166.
- Mullis,K., Faloona,F., Scharf,S., Saiki,R., Horn,G., and Erlich,H. (1986). Specific enzymatic amplification of DNA in vitro: the polymerase chain reaction. *Cold Spring Harb. Symp. Quant. Biol.* 51 Pt 1, 263-273.
- Murphy,G., Belda,F.J., Pau,C.P., Clewley,J.P., and Parry,J.V. (1999). Discrimination of subtype B and non-subtype B strains of human immunodeficiency virus type 1 by serotyping: correlation with genotyping. *J. Clin. Microbiol.* 37, 1356-1360.
- Neilson,J.R., John,G.C., Carr,J.K., Lewis,P., Kreiss,J.K., Jackson,S., Nduati,R.W., Mbori-Ngacha,D., Panteleeff,D.D., Bodrug,S., Giachetti,C., Bott,M.A., Richardson,B.A., Bwayo,J., Ndinya-Achola,J., and Overbaugh,J. (1999). Subtypes of human immunodeficiency virus type 1 and disease stage among women in Nairobi, Kenya. *J. Virol.* 73, 4393-4403.
- Neuman,d., V and Robinson,W.H. (2004). Microarray profiling of antiviral antibodies for the development of diagnostics, vaccines, and therapeutics. *Clin. Immunol.* 111, 196-201.
- O'Looney,N., Burgess,S.T., Chong,K.M., Ross,A.J., Robb,J., Forster,T., Beattie,J.S., Ghazal,P., Petrik,J., and Campbell,C.J. (2008). Evaluation of a protein microarray method for immunotyping erythrocytes in whole blood. *J. Immunoassay Immunochem.* 29, 197-209.
- Oberste,M.S. (2008). Comparative genomics of the coxsackie B viruses and related enteroviruses. *Curr. Top. Microbiol. Immunol.* 323, 33-47.
- Ofir,K., Berdichevsky,Y., Benhar,I., zriel-Rosenfeld,R., Lamed,R., Barak,Y., Bayer,E.A., and Morag,E. (2005). Versatile protein microarray based on carbohydrate-binding modules. *Proteomics.* 5, 1806-1814.
- Olson,K.B. (1999). Aum Shinrikyo: once and future threat? *Emerg. Infect. Dis.* 5, 513-516.
- Papa,A., Papadimitriou,E., Papoutsis,A., Kiosses,V., and Antoniadis,A. (2002). HIV-1 subtypes and circulating recombinant forms (CRFs) in Northern Greece. *Virus Res.* 85, 85-93.
- Parkinson,H., Kapushesky,M., Kolesnikov,N., Rustici,G., Shojatalab,M., Abeygunawardena,N., Berube,H., Dylag,M., Emam,I., Farne,A., Holloway,E., Lukk,M., Malone,J., Mani,R., Pilicheva,E., Rayner,T.F., Rezwan,F., Sharma,A., Williams,E., Bradley,X.Z., Adamusiak,T., Brandizi,M., Burdett,T., Coulson,R., Krestyaninova,M., Kurnosov,P., Maguire,E., Neogi,S.G., Rocca-Serra,P., Sansone,S.A., Sklyar,N., Zhao,M., Sarkans,U., and Brazma,A. (2009). ArrayExpress update--from an archive of functional genomics experiments to the atlas of gene expression. *Nucleic Acids Res.* 37, D868-D872.
- Parry,J. (2005). WHO launches "plan of war" to tackle avian flu. *BMJ* 331, 70.
- Pastinen,T., Kurg,A., Metspalu,A., Peltonen,L., and Syvanen,A.C. (1997). Minisequencing: a specific tool for DNA analysis and diagnostics on oligonucleotide arrays. *Genome Res.* 7, 606-614.
- Perez-Alvarez,L., Carmona,R., Munoz,M., Delgado,E., Thomson,M.M., Contreras,G., Pedreira,J.D., Rodriguez,R.R., Vazquez de,P.E., Medrano,L., Taboada,J.A., and Najera,R. (2003). High incidence of non-B and recombinant HIV-1 strains in newly diagnosed patients in Galicia, Spain: study of genotypic resistance. *Antivir. Ther.* 8, 355-360.

- Peterson,J.D., Umayam,L.A., Dickinson,T., Hickey,E.K., and White,O. (2001a). The Comprehensive Microbial Resource. *Nucleic Acids Res.* 29, 123-125.
- Peterson,J.D., Umayam,L.A., Dickinson,T., Hickey,E.K., and White,O. (2001b). The Comprehensive Microbial Resource. *Nucleic Acids Res.* 29, 123-125.
- Pham,T.T., Burchette,J.L., Jr., and Hale,L.P. (2003). Fatal disseminated adenovirus infections in immunocompromised patients. *Am. J. Clin. Pathol.* 120, 575-583.
- Plasterer,T.N. (1997). PRIMERSELECT. Primer and probe design. *Methods Mol. Biol.* 70, 291-302.
- Pope,M., Frankel,S.S., Mascola,J.R., Trkola,A., Isdell,F., Birx,D.L., Burke,D.S., Ho,D.D., and Moore,J.P. (1997). Human immunodeficiency virus type 1 strains of subtypes B and E replicate in cutaneous dendritic cell-T-cell mixtures without displaying subtype-specific tropism. *J. Virol.* 71, 8001-8007.
- Potter,C.W. (1967). The association of adenovirus infection with disease of the intestinal tract. *Mod. Trends Med. Virol.* 1, 162-181.
- Pyrce,K., Jebbink,M.F., Berkhout,B., and van der,H.L. (2008). Detection of new viruses by VIDISCA. Virus discovery based on cDNA-amplified fragment length polymorphism. *Methods Mol. Biol.* 454, 73-89.
- Qin,L., Rueda,L., Ali,A., and Ngom,A. (2005). Spot detection and image segmentation in DNA microarray data. *Appl. Bioinformatics.* 4, 1-11.
- Rajashekara,G., Eskra,L., Mathison,A., Petersen,E., Yu,Q., Harms,J., and Splitter,G. (2006). *Brucella*: functional genomics and host-pathogen interactions. *Anim Health Res. Rev.* 7, 1-11.
- Rajashekara,G., Glasner,J.D., Glover,D.A., and Splitter,G.A. (2004). Comparative whole-genome hybridization reveals genomic islands in *Brucella* species. *J. Bacteriol.* 186, 5040-5051.
- Ramisse,V., Patra,G., Garrigue,H., Guesdon,J.L., and Mock,M. (1996). Identification and characterization of *Bacillus anthracis* by multiplex PCR analysis of sequences on plasmids pXO1 and pXO2 and chromosomal DNA. *FEMS Microbiol. Lett.* 145, 9-16.
- Renjifo,B., Fawzi,W., Mwakagile,D., Hunter,D., Msamanga,G., Spiegelman,D., Garland,M., Kagoma,C., Kim,A., Chaplin,B., Hertzmark,E., and Essex,M. (2001). Differences in perinatal transmission among human immunodeficiency virus type 1 genotypes. *J. Hum. Virol.* 4, 16-25.
- Rice,P., Longden,I., and Bleasby,A. (2000). EMBOSS: the European Molecular Biology Open Software Suite. *Trends Genet.* 16, 276-277.
- Rico-Hesse,R. (2003). Microevolution and virulence of dengue viruses. *Adv. Virus Res.* 59, 315-341.
- Rimour,S., Hill,D., Militon,C., and Peyret,P. (2005). GoArrays: highly dynamic and efficient microarray probe design. *Bioinformatics.* 21, 1094-1103.
- Robertson,D.L., Anderson,J.P., Bradac,J.A., Carr,J.K., Foley,B., Funkhouser,R.K., Gao,F., Hahn,B.H., Kalish,M.L., Kuiken,C., Learn,G.H., Leitner,T., McCutchan,F., Osmanov,S., Peeters,M., Pieniazek,D., Salminen,M., Sharp,P.M., Wolinsky,S., and Korber,B. (2000). HIV-1 nomenclature proposal. *Science* 288, 55-56.
- Rotmistrovsky,K., Jang,W., and Schuler,G.D. (2004). A web server for performing electronic PCR. *Nucleic Acids Res.* 32, W108-W112.

- Ryan, M.A., Gray, G.C., Smith, B., McKeethan, J.A., Hawksworth, A.W., and Malasig, M.D. (2002). Large epidemic of respiratory illness due to adenovirus types 7 and 3 in healthy young adults. *Clin. Infect. Dis.* 34, 577-582.
- Sanchez, J.L., Binn, L.N., Innis, B.L., Reynolds, R.D., Lee, T., Mitchell-Raymundo, F., Craig, S.C., Marquez, J.P., Shepherd, G.A., Polyak, C.S., Conolly, J., and Kohlhasse, K.F. (2001). Epidemic of adenovirus-induced respiratory illness among US military recruits: epidemiologic and immunologic risk factors in healthy, young adults. *J. Med. Virol.* 65, 710-718.
- Sanchez-Seco, M.P., Hernandez, L., Eiros, J.M., Negredo, A., Fedele, G., and Tenorio, A. (2006). Detection and identification of orthopoxviruses using a generic nested PCR followed by sequencing. *Br. J. Biomed. Sci.* 63, 79-85.
- Sarantis, H., Johnson, G., Brown, M., Petric, M., and Tellier, R. (2004). Comprehensive detection and serotyping of human adenoviruses by PCR and sequencing. *J. Clin. Microbiol.* 42, 3963-3969.
- Schena, M., Shalon, D., Davis, R.W., and Brown, P.O. (1995). Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science* 270, 467-470.
- Schena, M., Shalon, D., Heller, R., Chai, A., Brown, P.O., and Davis, R.W. (1996). Parallel human genome analysis: microarray-based expression monitoring of 1000 genes. *Proc. Natl. Acad. Sci. U. S. A* 93, 10614-10619.
- Schmitz, H., Wigand, R., and Heinrich, W. (1983). Worldwide epidemiology of human adenovirus infections. *Am. J. Epidemiol.* 117, 455-466.
- Sellek, R., Jimenez, O., Aizpurua, C., Fernandez-Frutos, B., De, L.P., Camacho, M., Fernandez-Moreira, D., Ybarra, C., and Carlos, C.J. (2008). Recovery of *Francisella tularensis* from soil samples by filtration and detection by real-time PCR and cELISA. *J. Environ. Monit.* 10, 362-369.
- Shalon, D., Smith, S.J., and Brown, P.O. (1996). A DNA microarray system for analyzing complex DNA samples using two-color fluorescent probe hybridization. *Genome Res.* 6, 639-645.
- Shangkuan, Y.H., Chang, Y.H., Yang, J.F., Lin, H.C., and Shaio, M.F. (2001). Molecular characterization of *Bacillus anthracis* using multiplex PCR, ERIC-PCR and RAPD. *Lett. Appl. Microbiol.* 32, 139-145.
- Simon, R. and BRB-ArrayTools Development Team. BRB-ArrayTools. 2001.
Ref Type: Computer Program
- Southern, E.M., Maskos, U., and Elder, J.K. (1992). Analyzing and comparing nucleic acid sequences by hybridization to arrays of oligonucleotides: evaluation using experimental models. *Genomics* 13, 1008-1017.
- Spellman, P.T., Miller, M., Stewart, J., Troup, C., Sarkans, U., Chervitz, S., Bernhart, D., Sherlock, G., Ball, C., Lepage, M., Swiatek, M., Marks, W.L., Goncalves, J., Markel, S., Iordan, D., Shojatalab, M., Pizarro, A., White, J., Hubley, R., Deutsch, E., Senger, M., Aronow, B.J., Robinson, A., Bassett, D., Stoeckert, C.J., Jr., and Brazma, A. (2002). Design and implementation of microarray gene expression markup language (MAGE-ML). *Genome Biol.* 3, RESEARCH0046.
- Steel, L.F., Haab, B.B., and Hanash, S.M. (2005). Methods of comparative proteomic profiling for disease diagnostics. *J. Chromatogr. B Analyt. Technol. Biomed. Life Sci.* 815, 275-284.
- Stevens, W., Erasmus, L., Moloi, M., Taleng, T., and Sarang, S. (2008). Performance of a novel human immunodeficiency virus (HIV) type 1 total nucleic acid-based real-time PCR assay using whole blood and dried blood spots for diagnosis of HIV in infants. *J. Clin. Microbiol.* 46, 3941-3945.

- Syvanen,A.C. (2005). Toward genome-wide SNP genotyping. *Nat. Genet.* 37 *Suppl*, S5-10.
- Takeuchi,S., Itoh,N., Uchio,E., Aoki,K., and Ohno,S. (1999). Serotyping of adenoviruses on conjunctival scrapings by PCR and sequence analysis. *J. Clin. Microbiol.* 37, 1839-1845.
- Tarraga,J., Medina,I., Carbonell,J., Huerta-Cepas,J., Minguez,P., Alloza,E., Al-Shahrour,F., Vegas-Azcarate,S., Goetz,S., Escobar,P., Garcia-Garcia,F., Conesa,A., Montaner,D., and Dopazo,J. (2008). GEPAS, a web-based tool for microarray data analysis and interpretation. *Nucleic Acids Res.* 36, W308-W314.
- Thibault,F.M., Valade,E., and Vidal,D.R. (2004). Identification and discrimination of *Burkholderia pseudomallei*, *B. mallei*, and *B. thailandensis* by real-time PCR targeting type III secretion system genes. *J. Clin. Microbiol.* 42, 5871-5874.
- Thomas,R., Johansson,A., Neeson,B., Isherwood,K., Sjostedt,A., Ellis,J., and Titball,R.W. (2003). Discrimination of human pathogenic subspecies of *Francisella tularensis* by using restriction fragment length polymorphism. *J. Clin. Microbiol.* 41, 50-57.
- Thompson,J.D., Gibson,T.J., and Higgins,D.G. (2002). Multiple sequence alignment using ClustalW and ClustalX. *Curr. Protoc. Bioinformatics.* Chapter 2, Unit.
- Titball,R.W. and Petrosino,J.F. (2007). *Francisella tularensis* genomics and proteomics. *Ann. N. Y. Acad. Sci.* 1105, 98-121.
- Touchman,J.W., Wagner,D.M., Hao,J., Mastrian,S.D., Shah,M.K., Vogler,A.J., Allender,C.J., Clark,E.A., Benitez,D.S., Youngkin,D.J., Girard,J.M., Auerbach,R.K., Beckstrom-Sternberg,S.M., and Keim,P. (2007). A North American *Yersinia pestis* draft genome sequence: SNPs and phylogenetic analysis. *PLoS. ONE.* 2, e220.
- Tucker,J.B. (1999). Historical trends related to bioterrorism: An empirical analysis. *Emerg. Infect. Dis.* 5, 498-504.
- van,d., V (1963). THE ROLE OF ADENOVIRUSES IN RESPIRATORY DISEASE. *Am. Rev. Respir. Dis.* 88, SUPPL-80.
- van,d., V and DIJKMAN,J.H. (1962). Association of type 21 adenovirus with acute respiratory illness in military recruits. *Am. J. Hyg.* 76, 149-159.
- van,d., V, Oei,K.G., and Abarbanel,M.F. (1969). Patterns of infections with adenovirus types 4, 7 and 21 in military recruits during a 9-year survey. *J. Hyg. (Lond)* 67, 255-268.
- Venter,J.C., Adams,M.D., Myers,E.W., Li,P.W., Mural,R.J., Sutton,G.G., Smith,H.O., Yandell,M., Evans,C.A., Holt,R.A., Gocayne,J.D., Amanatides,P., Ballew,R.M., Huson,D.H., Wortman,J.R., Zhang,Q., Kodira,C.D., Zheng,X.H., Chen,L., Skupski,M., Subramanian,G., Thomas,P.D., Zhang,J., Gabor Miklos,G.L., Nelson,C., Broder,S., Clark,A.G., Nadeau,J., McKusick,V.A., Zinder,N., Levine,A.J., Roberts,R.J., Simon,M., Slayman,C., Hunkapiller,M., Bolanos,R., Delcher,A., Dew,I., Fasulo,D., Flanigan,M., Florea,L., Halpern,A., Hannenhalli,S., Kravitz,S., Levy,S., Mobarry,C., Reinert,K., Remington,K., bu-Threideh,J., Beasley,E., Biddick,K., Bonazzi,V., Brandon,R., Cargill,M., Chandramouliswaran,I., Charlab,R., Chaturvedi,K., Deng,Z., Di,F., V, Dunn,P., Eilbeck,K., Evangelista,C., Gabrielian,A.E., Gan,W., Ge,W., Gong,F., Gu,Z., Guan,P., Heiman,T.J., Higgins,M.E., Ji,R.R., Ke,Z., Ketchum,K.A., Lai,Z., Lei,Y., Li,Z., Li,J., Liang,Y., Lin,X., Lu,F., Merkulov,G.V., Milshina,N., Moore,H.M., Naik,A.K., Narayan,V.A., Neelam,B., Nusskern,D., Rusch,D.B., Salzberg,S., Shao,W., Shue,B., Sun,J., Wang,Z., Wang,A., Wang,X., Wang,J., Wei,M., Wides,R., Xiao,C., Yan,C., Yao,A., Ye,J., Zhan,M., Zhang,W., Zhang,H., Zhao,Q., Zheng,L., Zhong,F., Zhong,W., Zhu,S., Zhao,S., Gilbert,D., Baumhueter,S., Spier,G., Carter,C., Cravchik,A., Woodage,T., Ali,F., An,H., Awe,A., Baldwin,D., Baden,H., Barnstead,M., Barrow,I., Beeson,K., Busam,D., Carver,A., Center,A., Cheng,M.L., Curry,L., Danaher,S., Davenport,L., Desilets,R., Dietz,S., Dodson,K., Doup,L., Ferriera,S., Garg,N., Gluecksmann,A., Hart,B., Haynes,J., Haynes,C., Heiner,C., Hladun,S., Hostin,D., Houck,J., Howland,T., Ibegwam,C., Johnson,J., Kalush,F., Kline,L., Koduru,S., Love,A.,

Mann,F., May,D., McCawley,S., McIntosh,T., McMullen,I., Moy,M., Moy,L., Murphy,B., Nelson,K., Pfannkoch,C., Pratts,E., Puri,V., Qureshi,H., Reardon,M., Rodriguez,R., Rogers,Y.H., Romblad,D., Ruhfel,B., Scott,R., Sitter,C., Smallwood,M., Stewart,E., Strong,R., Suh,E., Thomas,R., Tint,N.N., Tse,S., Vech,C., Wang,G., Wetter,J., Williams,S., Williams,M., Windsor,S., Winn-Deen,E., Wolfe,K., Zaveri,J., Zaveri,K., Abril,J.F., Guigo,R., Campbell,M.J., Sjolander,K.V., Karlak,B., Kejariwal,A., Mi,H., Lazareva,B., Hatton,T., Narechania,A., Diemer,K., Muruganujan,A., Guo,N., Sato,S., Bafna,V., Istrail,S., Lippert,R., Schwartz,R., Walenz,B., Yooseph,S., Allen,D., Basu,A., Baxendale,J., Blick,L., Caminha,M., Carnes-Stine,J., Caulk,P., Chiang,Y.H., Coyne,M., Dahlke,C., Mays,A., Dombroski,M., Donnelly,M., Ely,D., Esparham,S., Fosler,C., Gire,H., Glanowski,S., Glasser,K., Glodek,A., Gorokhov,M., Graham,K., Gropman,B., Harris,M., Heil,J., Henderson,S., Hoover,J., Jennings,D., Jordan,C., Jordan,J., Kasha,J., Kagan,L., Kraft,C., Levitsky,A., Lewis,M., Liu,X., Lopez,J., Ma,D., Majoros,W., McDaniel,J., Murphy,S., Newman,M., Nguyen,T., Nguyen,N., and Nodell,M. (2001). The sequence of the human genome. *Science* 291, 1304-1351.

Vergne,L., Bourgeois,A., Mpoudi-Ngole,E., Mougnotou,R., Mbuagbaw,J., Liegeois,F., Laurent,C., Butel,C., Zekeng,L., Delaporte,E., and Peeters,M. (2003). Biological and genetic characteristics of HIV infections in Cameroon reveals dual group M and O infections and a correlation between SI-inducing phenotype of the predominant CRF02_AG variant and disease stage. *Virology* 310, 254-266.

Walter,B.L., Armitage,A.E., Graham,S.C., de,O.T., Skinhoj,P., Jones,E.Y., Stuart,D.I., McMichael,A.J., Chesebro,B., and Iversen,A.K. (2009). Functional characteristics of HIV-1 subtype C compatible with increased heterosexual transmissibility. *AIDS*.

Wang,D., Urisman,A., Liu,Y.T., Springer,M., Ksiazek,T.G., Erdman,D.D., Mardis,E.R., Hickenbotham,M., Magrini,V., Eldred,J., Latreille,J.P., Wilson,R.K., Ganem,D., and DeRisi,J.L. (2003). Viral discovery and sequence recovery using DNA microarrays. *PLoS. Biol.* 1, E2.

Wernersson,R. and Nielsen,H.B. (2005). OligoWiz 2.0--integrating sequence feature annotation into the design of microarray probes. *Nucleic Acids Res.* 33, W611-W615.

Whetzel,P.L., Parkinson,H., Causton,H.C., Fan,L., Fostel,J., Fragoso,G., Game,L., Heiskanen,M., Morrison,N., Rocca-Serra,P., Sansone,S.A., Taylor,C., White,J., and Stoeckert,C.J., Jr. (2006). The MGED Ontology: a resource for semantics-based description of microarray experiments. *Bioinformatics.* 22, 866-873.

Williams,J.C., Case-Green,S.C., Mir,K.U., and Southern,E.M. (1994). Studies of oligonucleotide interactions by hybridisation to arrays: the influence of dangling ends on duplex yield. *Nucleic Acids Res.* 22, 1365-1367.

Wilson,W.J., Strout,C.L., DeSantis,T.Z., Stilwell,J.L., Carrano,A.V., and Andersen,G.L. (2002). Sequence-specific identification of 18 pathogenic microorganisms using microarray technology. *Mol. Cell Probes* 16, 119-127.

Wingren,C. and Borrebaeck,C.A. (2008). Antibody microarray analysis of directly labelled complex proteomes. *Curr. Opin. Biotechnol.* 19, 55-61.

Xing,J., Watkins,W.S., Witherspoon,D.J., Zhang,Y., Guthery,S.L., Thara,R., Mowry,B.J., Bulayeva,K., Weiss,R.B., and Jorde,L.B. (2009). Fine-scaled human genetic structure revealed by SNP microarrays. *Genome Res.* 19, 815-825.

Yin,J.Q., Zhao,R.C., and Morris,K.V. (2008). Profiling microRNA expression with microarrays. *Trends Biotechnol.* 26, 70-76.

Zhu,H., Hu,S., Jona,G., Zhu,X., Kreiswirth,N., Willey,B.M., Mazzulli,T., Liu,G., Song,Q., Chen,P., Cameron,M., Tyler,A., Wang,J., Wen,J., Chen,W., Compton,S., and Snyder,M. (2006). Severe acute respiratory syndrome diagnostics using a coronavirus protein microarray. *Proc. Natl. Acad. Sci. U. S. A* 103, 4011-4016.

9. Anexo I – **Publicaciones**

A lo largo de este trabajo se han generado las siguientes publicaciones:

Lopez-Campos,G., Coiras,M., Sanchez-Merino,J.P., Lopez-Huertas,M.R., Spiteri,I., Martin-Sanchez,F., and Perez-Brena,P. (2007). Oligonucleotide microarray design for detection and serotyping of human respiratory adenoviruses by using a virtual amplicon retrieval software. *J. Virol. Methods* *145*, 127-136.

Lopez-Campos,G., Lopez,A., V, and Martin-Sanchez,F. (2008). Addressing the biomedical informatics needs of a microarray laboratory in a clinical microbiology context. *Stud. Health Technol. Inform.* *136*, 45-50.

Lopez-Campos,G.H., Garcia-Albert,L., Martin-Sanchez,F., and Garcia-Saiz,A. (2006). Analysis and management of HIV peptide microarray experiments. *Methods Inf. Med.* *45*, 158-162.

Analysis and Management of HIV Peptide Microarray Experiments

G. H. López-Campos¹, L. García-Albert², F. Martín-Sánchez¹, A. García-Sáiz²

¹Medical Bioinformatics Department, Institute of Health "Carlos III", Majadahonda, Madrid, Spain

²Retrovirus and Papillomavirus Reference and Diagnostics Department, Institute of Health "Carlos III", Majadahonda, Madrid, Spain

Summary

Objectives: To develop a tool for then easy and user-friendly management of peptide microarray experiments and for the use of the results of these experiments for the study the immune response against HIV virus infection in clinical samples.

Methods: Applying bioinformatics and statistics for the analysis of data coming from microarray experiments as well as implementing a MIAME (Minimum Information About a Microarray Experiment) compliant database for managing and annotating experiments, results and samples.

Results: We present a new tool for managing not only nucleic acid microarray experiments but also protein microarray experiments. From the analysis of experimental data, we can detect different profiles in the reactivity of the sera with different genotypes.

Conclusions: We have developed a new tool for managing microarray data including clinical annotations for the samples as well as the capability of annotating other microarray formats different to those based on nucleic acids. The use of peptide microarrays and bioinformatics analysis opens a new scope for the characterization of the immune response, and analyzing and identifying the humoral response of viruses with different genotypes.

Keywords

Protein array analysis, HIV, serotyping, HIV infection, computational biology

Methods Inf Med 2006; 45: 158–26

1. Introduction

Human Immunodeficiency Virus (HIV) is one of the most prominent public health problems today. Even with the study of HIV virus having generated an amazing increase in the knowledge of the biology and pathogenesis of the virus in the last 20 years, there are still several aspects that remain unclear. The greater and deeper understanding of the mechanisms related with viral infection and the response of the immune system is a key issue for fighting against the disease and its spread in modern societies.

HIV virus can be classified in different groups and subtypes depending on the phylogenetic relationships of their nucleic acid sequences. The study and analysis of these subtypes [1, 2] is an important tool for analyzing the behavior of the virus and the patterns of its dissemination as well as for increasing the information related with this infection [3, 4]. HIV virus presents a high genetic heterogeneity and traditionally has been studied from different points of view combining techniques based on the analysis of both nucleic acids and proteins. One of the most commonly used assays for HIV detection and characterization are ELISA tests. These tests are commonly used as routine assays for analyzing the humoral response of infected individuals against the virus.

Microarrays have been one of the hottest topics in biology and bioinformatics in the last years. The immobilization of proteins [5], antibodies [6] and peptides has allowed the possibility of carrying over ELISA-like tests in a microarray format [7]. While they originally focused on gene expression analysis [8–10], nowadays the possibility of immobilizing new molecules to their sur-

faces is extending their application into different fields and opening new scopes for their clinical application.

The management and analysis of the huge amounts of data and information generated during all the processes done during microarray experiments (from manufacturing to the final stages of data collection and analysis) requires the application and development of bioinformatics tools. Within this work we apply all these techniques for the analysis of the humoral response against HIV infection, using peptide microarrays and identifying the reactivity profiles of infected sera coming from different virus genotypes.

2. Objectives

The aim of this work is the study of the humoral response during HIV infection, applying peptide microarrays and bioinformatics tools for the analysis and management of data generated during the experiments.

The use of "in-house" built peptide microarrays makes the storage and annotation of information related to the manufacturing process and the probes immobilized on the microarrays as well as with the management of information related with the samples and the assays necessary. The development of a database with LIMS (Laboratory Information Management System) capabilities is therefore necessary, allowing the inclusion of clinical annotation of the samples and following the current available standard MIAME (Minimum Information About a Microarray Experiment) [11] to facilitate the exchange of information.

The analysis of the data from microarray experiments for the characterization of the immune response during HIV infection is also a major goal of the study. We explore the capacity of peptide microarrays to identify the antibody response against HIV peptides from different genotypes of the virus.

3. Methods

The selection of samples used for the study was done based on previously published data [12]. Finally 41 sera samples from different subtypes of HIV positive and negative patients were selected for undergoing an ELISA-like protein microarray study (Table 1).

Protein microarray manufacture and design has already been presented (L. Garcia-Albert et al. Spanish VII National Virology Congress Barcelona, Oct. 2003). Each microarray is built up using a four-point dilution series of eight peptides that will allow the analysis of the pattern of the reactivity. These sera were incubated with the protein arrays and afterwards reacted with a Cy-3 labeled antihuman antibody acting as conjugate.

Image 4.0 (Biodiscovery Inc., Los Angeles, CA, USA) was used for image quantification. The algorithm applied for image quantification is based on the histogram analysis of the images. Signal and local background intensities were calculated and for each spot, the background subtracted signal was used for further analysis.

As a quality control measure within the slide, an ANOVA test was applied to check the reproducibility of the printing process and the responses of the peptides bound to the surface. Also, a second ANOVA test was applied to analyze the behavior of the technical replicates used to validate the laboratory technique. Only the slides that passed these tests were used for data analysis. From a total of 54 microarray assays (including technical replicates done), 52 passed the test and only two didn't fulfill the ANOVA test for quality control.

Applying a normalization algorithm is necessary in order to compare the results within different slides. For this purpose we

Table 1 Number of sera of each of the genotypes analyzed and used for the microarray experiments

Sera genotype	Number of different sera used
A	3
B	10
HIV-positive B-sera but ELISA negative	4
C	3
D	3
E	3
F	3
G	2
O	2
HIV-negative	8

used the signals coming from IgG90 spots and we calculated a normalization index that equaled the IgG90 signal to 1 in all the slides, then the values were transformed calculating their log 2. These log-transformed and normalized values were used for the data analysis. The final number of data points analyzed coming from each microarray was 34. The analysis of data was done using BRB Array Tools software (<http://linus.nci.nih.gov/BRB-ArrayTools.html>). BRB Array Tools is an R-based suite for microarray data analysis with an MS ExcelTM-based interface with a complete set of tools and algorithms that facilitates the analysis of data.

4. Results

The design and development of a database has allowed us to store all the information related to the experiments. The system has improved the data retrieval process for the printing and manufacturing stages as well as for the analysis of the results coming from the experiments.

AMANDA (Array MANagement Database) has been designed in compliance with the current MIAME standard and as a tool for tracking all the aspects related to microarray processing. The database has been designed to allow the extension of MIAME to include clinical annotation of samples, as well as to annotate other probes different

from nucleic acids, like the peptides used in this work. The database was designed using a relational model under MS AccessTM. MS Access was selected because it is a tool that is widely distributed and easy to use and maintain for people who are working in a laboratory. It doesn't require computational skills or powerful computers. AMANDA's interface is designed to offer two functionalities; one is designed for data entry and management as administrator, using several forms. The second functionality, offered to users only, involves information retrieval by querying the database. Some tools have been integrated into the design of AMANDA with the aim of facilitating data input and output, also including the tracking of the processes done during data analysis.

The data stored in AMANDA were used for the analysis of the immune response during HIV infection. The first and most intuitive comparison is based on a scatter plot of the averages of positive and negative sera reactivities against the antigens immobilized on the microarray surface (Fig. 1). Using this first analysis it is possible to detect that all the spots have greater intensities in the infected samples. It is also possible to appreciate the linearity in the responses of the different dilution points used in the array. The scatter plot analysis of the sera shows that there are some antigens with a differential reactivity between positive and negative sera.

The BRB Array Tools software was used for a statistical analysis of the data. The comparison and characterization of the statistically significant differences between both positive and negative sera was analyzed using a BRB Array tools class comparison tool. The analyses were always done using a nominal significance for the univariate test of 0.01, a maximum number of allowed false-positive calls to the antigens of 0 and on the basis of 1000 permutations for the permutation test. From this analysis it is possible to identify six antigens significant at the 0.01 level of the univariate test, having a probability of getting at least six genes significant by chance (at the 0.01 level) if there are no real differences between the classes of 0.015. It's also possible to expand this list up to 16 antigens on the basis of the number of allowed false-positives.

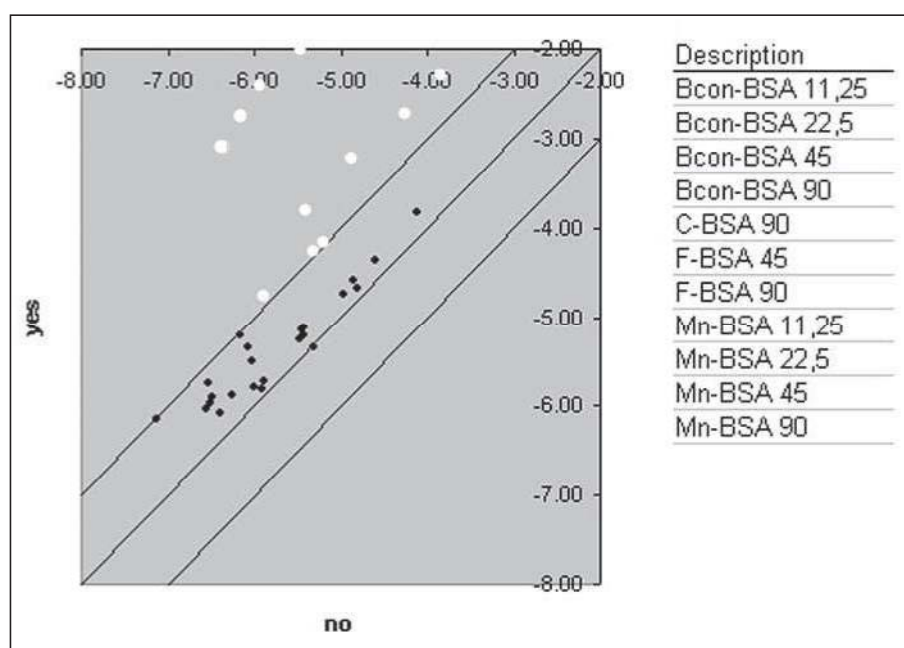


Fig. 1 Scatter plot for the comparison of the average normalized log 2 intensities from HIV-positive vs HIV-negative sera. The lines are set at a twofold difference and the spots highlighted in white have a difference in intensity above twofold. It is also possible to appreciate the linearity in the response of the dilution series.

tives (0). The results of this analysis are shown in Table 2.

The same analysis was carried out to characterize each of the genotypes of the sera available for this study, comparing the profiles with the aim of finding which

peptides best characterize each of the genotypes. In this case the same parameters as in the previous analyses were used and the results are shown in Table 3. It is possible to find a “finger-print” for five of eight subtype genotypes.

Description	Parametric p-value	Permutation p-value
BMn 45	0.00019	0
BMn 90	0.00026	2e-04
BMn 22,5	0.00043	4e-04
BMn 11,25	0.00141	0.0011
F 45	0.00630	0.0066
Bcon 22,5	0.00777	0.0074
Bcon 90	0.01060	0.0092
Bcon 11,25	0.01261	0.0120
C 11,25	0.01351	0.0123
Bcon 45	0.01593	0.0145
E 11,25	0.02002	0.0183
F 90	0.02156	0.0177
C 90	0.02189	0.0182
C 45	0.03027	0.0238
F 22,5	0.03144	0.0268
F 11,25	0.05336	0.0482

Table 2 Identification of the statistically significant antigens detected for discriminating positive and negative sera. The first six antigens are the ones positive at the 0.01 level. All the others are selected because they fulfill the other parameters of the analysis (0 false discoveries and 0.01% of false discoveries).

5. Discussion

We have designed AMANDA, a database that allows us to track information of all the steps related to the executed experiments, from printing to the final steps of normalization and data analysis. The designed database is a MIAME-compliant tool with some extended capabilities that allows its use as a laboratory information management system. The use of this database has supported the analysis of the data providing some analytical tools embedded in it and has also served as the main data repository for the analysis and storage of all the information generated during the experiments.

The use of peptide microarrays and bioinformatics analysis using class comparison techniques has already successfully been applied in the analysis of the immune response in vaccine trials [13]. We have also applied a similar approach based in microarray experiments and class comparison analysis in the analysis of the response against different subtypes of HIV virus. The analysis of immobilized dilution series of the same antigens allows us to establish a more precise reactivity profile between the eight different peptides immobilized on the array surface. The selection of the significance level at 0.01 was done because, in contrast with the stringent conditions applied in gene expression studies where there are thousands of genes and data points, we are working only with 34 data points. This reduced number of data points allows us to decrease the significance level of the analysis.

From this study we propose that it is possible to identify what the peptides and dilutions are that best discriminate between positive and negative sera giving a statistically significant increase in the intensities of the reaction signals. The list of significant reactive peptides can be extended (Table 2), identifying more peptides that discriminate between both kinds of sera. This extended list is more similar to the results obtained from the first and more visual scatter plot analysis based on the average intensity of the positive vs. negative sera.

After identifying which peptides are the ones that best characterize the differences between positive and negative sera, the next

step in the analysis was to analyze the response against the different genotypes of the viruses. This analysis was done comparing the reactivity profiles of sera from the different genotypes with all the other sera. From this analysis we have been able to identify the patterns of reactivity against the immobilized peptides that characterize the antibody response against the different genotypes. The peptides immobilized on the array came from consensus sequences of the different genotypes of the virus and should be able to identify the different genotypes of the viruses. From our analysis, we have seen that this does not always happen. These differences between genotypes and serotypes have been previously reported [14]. An interesting example of this effect can be observed in the B-positive sera, where the discriminating feature is a decrease in the intensity of the reaction against the dilutions of peptide A. This can be explained because as we have previously seen, the BMn and Bcon peptides are discriminators for all positive sera. We can also detect discriminating peptides and dilutions by having a significant increase in the intensity of the reactions for sera of the subtypes D, E. Sera from G genotype virus share the discriminating dilutions of peptide E with E genotype samples, but in this case the discrimination is due to a statistically significant decrease in the intensity of the reactions against this dilutions of peptide E. There were three cases, sera with virus of genotypes A, C and F, where it has not been possible to identify a set of antigens which have a statistically different reactivity across all the other sera under the conditions of the analysis. A more detailed analysis of these sera (data not shown) shows that genotypes A and C share the same reactivity profile and this profile partially overlaps with the profile of E sera.

The approach based on the analysis of reactivity profiles using the response against several peptides at once might be used as a tool for predictive models that could be useful in research and future diagnostics. We have explored the possibility of identifying classifiers for the prediction of positive and negative sera (data not shown) finding models that correctly classify up to 80% of the samples. We have observed that

Table 3 The table shows the peptides and dilutions found to be statistically significant to discriminate among the different genotypes of HIV. Peptides were selected from V3 region of gp120 viral protein of each viral subtype except for the B subtype where two peptides, BMn and Bcon, were selected. It is possible to also appreciate the differences in the intensities that allow the discrimination.

Genotype	Peptide	Parametric F-test p-value	Permutation F-test p-value	x-fold difference in the intensity
O	O 90	8.20E-14	0.001	12
	O 45	1.15E-11	0.001	10.56
A	No significant peptides detected			
B	A 11,25	0.00089	0.001	-2.33
	A 22,5	0.00139	0.0015	-2.44
	A 45	0.00262	0.0015	-2.67
	A 90	0.0054	0.0065	-2.35
C	No significant peptides detected			
D	D 90	3.11E-07	0.0015	5.44
	D 11,25	3.19E-07	0.0015	5.76
	D 45	1.21E-06	0.0035	4.94
	D 22,5	1.74E-06	0.004	4.26
E	C 45	0.00022	0.001	5.66
	C 22,5	0.000336	0.0015	5.23
	C 90	0.000598	0.001	5.26
	C 11,25	0.000638	0.0025	2.18
	E 45	0.000673	0.0025	1.91
	E 90	0.00211	0.006	1.91
E	E 11,25	0.00392	0.004	2.33
F	No significant peptides detected			
G	E 22,5	0.00019	0.001	-2.5
	E 45	0.00495	0.0085	-2.3
	E 90	0.00845	0.012	-2

some of the misclassifications could be explained by the low reactivity of the sera, accordingly with data coming from ELISA tests.

6. Conclusions

New high throughput technologies such as microarrays can be used for analyzing complex humoral responses such as those produced during HIV infections, but they require bioinformatics support for data management and analysis. The huge amounts of data generated during these experiments as well as the need to have good annotation of all the samples and processes carried out have fostered the design of a new

tool, “AMANDA”, that is able to manage all of this information. The possibility of using it in small laboratories using common PCs and software is important because it facilitates its use in laboratories without any previous special computational requirements. The use of this tool in this study has improved not only the storage but also the analysis of data from the experiments. The information stored and analyzed has allowed us to identify the peptide that best characterizes the response against HIV infection in the tested sera. The study of sera infected with different genotype viruses allowed us to identify the peptides that offer the best discriminating response among the different genotypes. The use of these approaches may open a new way for the analy-

sis of a complex problem as is the process of subtyping HIV on the basis of the immune response where several other approaches have been tried. The possibility of using profiles and the simultaneous analysis of the reactivity against several peptides in a single reaction can be applied in the detection of discriminating profiles that can be used as predictors to analyze whether a sample is infected or not. We have explored this possibility but, due to the number of samples available for the study, we can only offer the results as an exploratory result.

Microarrays are a very powerful tool that can be applied for the analysis of the immune responses. The possibility of using different supervised classification methods combined with the data coming from a high throughput technique such as microarrays is opening new avenues for research.

The deep detail of these types of interactions may help and improve our knowledge about how our immune systems fight and help us to protect ourselves against pathogenic infections. The development of new analytical techniques combined with massive experimental data is offering a new scope for improving and refining current diagnostic tools and it might also become a powerful tool for the analysis of more complex traits in the future.

Acknowledgment

Analyses were performed using BRB Array-Tools v3.1 developed by Dr. R Simon and Amy Peng Lam. We would like also to thank O. García and I. Hermosilla for their kind help reading the manuscript.

References

- Robertson DL, Anderson JP, Bradac JA, Carr JK, Foley B, Funkhouser RK, et al. HIV-1 nomenclature proposal. *Science* 2000; 288 (5463): 55-6.
- Laukkanen T, Albert J, Liitsola K, Green SD, Carr JK, Leitner T, et al. Virtually full-length sequences of HIV type 1 subtype J reference strains. *AIDS Res Hum Retroviruses* 1999; 15 (3): 293-7.
- Soto-Ramirez LE, Renjifo B, McLane MF, Marlink R, O'Hara C, Sutthent R, et al. HIV-1 Langerhans' cell tropism associated with heterosexual transmission of HIV. *Science* 1996; 271 (5253): 1291-3.
- Essex M, Soto-Ramirez LE, Renjifo E, Wang WK, Lee TH. Genetic variation within human immunodeficiency viruses generates rapid changes in tropism, virulence, and transmission. *Leukemia* 1997; 11 (Suppl 3): 93-4.
- MacBeath G, Schreiber SL. Printing proteins as microarrays for high-throughput function determination. *Science* 2000; 289 (5485): 1760-3.
- Angenendt P, Glokler J, Murphy D, Lehrach H, Cahill DJ. Toward optimized antibody microarrays: a comparison of current microarray support materials. *Analytical Biochemistry* 2002; 309 (2): 253-60.
- Joos TO, Schrenk M, Hopfl P, Kroger K, Chowdhury U, Stoll D, et al. A microarray enzyme-linked immunosorbent assay for autoimmune diagnostics. *Electrophoresis* 2000; 21 (13): 2641-50.
- Schena M, Shalon D, Davis RW, Brown PO. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science* 1995; 270 (5235): 467-70.
- Schena M, Shalon D, Heller R, Chai A, Brown PO, Davis RW. Parallel human genome analysis: microarray-based expression monitoring of 1000 genes. *Proc Natl Acad Sci USA* 1996; 93 (20): 10614-9.
- DeRisi JL, Iyer VR, Brown PO. Exploring the metabolic and genetic control of gene expression on a genomic scale. *Science* 1997; 278 (5338): 680-6.
- Brazma A, Hingamp P, Quackenbush J, Sherlock G, Spellman P, Stoeckert C, et al. Minimum information about a microarray experiment (MIAME) – toward standards for microarray data. *Nat Genet* 2001; 29 (4): 365-71.
- Garcia-Albert L, Ortiz M, Garcia-Saiz A. HIV type 1 non-B subtype prevalence in Spain, 1997-1998. *AIDS Res Hum Retroviruses* 2001; 17 (14): 1317-20.
- Neuman de Vegvar HE, Amara RR, Steinman L, Utz PJ, Robinson HL, Robinson WH. Microarray Profiling of Antibody Responses against Simian-Human Immunodeficiency Virus: Postchallenge Convergence of Reactivities Independent of Host Histocompatibility Type and Vaccine Regimen. *J Virol* 2003; 77 (20): 11125-38.
- Murphy G, Belda FJ, Pau CP, Clewley JP, Parry JV. Discrimination of subtype B and non-subtype B strains of human immunodeficiency virus type 1 by serotyping: correlation with genotyping. *J Clin Microbiol* 1999; 37 (5): 1356-60.

Correspondence to:

Guillermo López-Campos
Institute of Health "Carlos III"
Ctra. Majadahonda a Pozuelo, km 2
28220 Majadahonda
Madrid
Spain
E-mail: glopez@isciii.es

Oligonucleotide microarray design for detection and serotyping of human respiratory adenoviruses by using a virtual amplicon retrieval software

Guillermo López-Campos^{a,*}, Mayte Coiras^b, Juan Pedro Sánchez-Merino^a, María Rosa López-Huertas^b, Inmaculada Spiteri^a, Fernando Martín-Sánchez^a, Pilar Pérez-Breña^b

^a Medical Bioinformatics Unit, National Center of Microbiology, Instituto de Salud Carlos III, Ctra. Majadahonda-Pozuelo, km 2, 28220 Majadahonda, Madrid, Spain

^b Virology Service, National Center of Microbiology, Instituto de Salud Carlos III, Ctra. Majadahonda-Pozuelo, km 2, 28220 Majadahonda, Madrid, Spain

Received 15 January 2007; received in revised form 10 May 2007; accepted 16 May 2007
Available online 21 June 2007

Abstract

The rapid development of new molecular biology methods has improved infectious disease diagnosis, which is increasingly important to clinical management and public health. A wide variety of new methods which are more specific, sensitive and robust, such as combination of PCR and microarray technology, has gradually replaced the conventional assays usually used in routine diagnosis. Both methods have the advantage of speed and sensitivity but tend to be expensive and technically demanding. Therefore, it is necessary to develop more simple assays that could be available for all diagnostic laboratories. To this aim, a simple microarray assay for detection and typing of adenoviruses causing acute respiratory disease in humans was developed. The absence of effective therapeutic or alternative prophylactic treatment for this infection makes essential its rapid diagnosis to implement fast control measures. Moreover, the family Adenoviridae includes numerous serotype groups and constitutes an ideal model system to develop diagnostic methods for other human pathogens. For their precise identification, an amplicon retrieval software that simplifies and accelerates the processing of all sequences necessary to perform this type of experiments has also been developed. This new technology was applied for the simultaneous detection and serotyping of acute respiratory disease-causing adenoviruses in laboratory and clinical samples with efficiency and accuracy.

© 2007 Elsevier B.V. All rights reserved.

Keywords: Microarray; Probe design; Amplicon retrieval software; Adenovirus; Acute respiratory disease

1. Introduction

Human adenoviruses are a family of double-stranded DNA viruses that are classified into 52 different serotypes (De Jong et al., 1999; Shenk, 2001) based on the neutralization infectivity assays (Shenk, 2001). According to their biological and physicochemical properties and their nucleic acid homologies, the different serotypes of adenoviruses have been further divided into six species (A–F) (Schmitz et al., 1983; Shenk, 2001). Distinction of adenovirus serotypes is based on the neutralization of infectivity, which is directed against epitopes on the hexon protein, and hemagglutination inhibition (Willcox and Mautner, 1976). Adenovirus infection causes a broad spectrum of dis-

eases that can affect the gastrointestinal, ocular and respiratory tract systems (Horwitz, 2001; Shenk, 2001). In particular, the adenoviruses that are mainly isolated from patients suffering respiratory tract diseases are serotypes 1, 2, 3, 4, 5, 6 and 7 (Schmitz et al., 1983). Primarily, the disease produced by adenovirus 7a in children is almost exclusive of the respiratory tract, and in some individuals it may be very severe and possibly fatal (Mitchell et al., 2000). Other adenoviruses involved in respiratory infection are types 4, 14 and 21 (Faix et al., 2004; Van der Veen and Dijkman, 1962). All these viruses spread rapidly in closed environments often causing epidemic disease in crowded communities. On the other hand, adenovirus infection is also difficult to distinguish clinically from other viral or bacterial respiratory infections (Horwitz, 2001). All these factors together indicate that clinical microbiological methods are necessary to detect the acute respiratory disease caused by adenoviruses and to determine the serotypes involved. Nowadays, the possibility

* Corresponding author. Tel.: +34 91 8223246; fax: +34 91 5097919.
E-mail address: glopez@isci.es (G. López-Campos).

of developing antiviral therapy or prophylaxis for some respiratory viral infections (influenza, respiratory syncytial virus or parainfluenza) seems high, but is still extremely low for adenovirus acute respiratory disease. Rapid diagnostic methods could allow to avoid inadequate antiviral treatment and to focus in the control of viral transmission of adenovirus, especially in the nosocomial infection. Such a method would be necessary to aid in controlling viral transmission and adenovirus-associated respiratory disease (Grondahl et al., 1999; Poddar, 1999). In summary, efficient laboratory diagnosis of infectious diseases caused by adenoviruses is therefore increasingly important to clinical management and public health.

Conventional methods for adenovirus detection and serotyping involve time-consuming and labour-intensive methods such as viral shell culture, observation for cytopathic effects, immunofluorescence assays, electron microscopy, microneutralization assays or serotyping. These methods produce results in 3 days to 3 weeks and as a consequence, the acquisition of results generally does not mean any relevance for the treatment or quarantine of infected individuals. In order to address these deficits, current molecular techniques to directly detect nucleic acids of microbial pathogens in clinical specimens such as polymerase chain reaction (PCR)-based assays have been recently employed in clinical diagnosis of adenoviruses (Avellón et al., 2001; Faix et al., 2004; Na et al., 2002; Papapetropoulou and Vantarakis, 1998; Pring-Akerblom et al., 1999; Xu and Erdman, 2001; Xu et al., 2000). These molecular techniques are rapid, sensitive, and may succeed when culturing the organism fails as well as allow quick control of the outbreaks. Recently, other methods such as PCR plus sequencing (Takeuchi et al., 1999), or a combination of PCR and restriction endonuclease analysis have been developed to serotype adenovirus (Allard et al., 2001; Saitoh-Inagawa et al., 1996), as well as the microarray technology (Lin et al., 2004). However, the use of PCR in molecular serotypes differentiation has several limitations due to the relatively small number of serotypes detectable in a single assay (Broude et al., 2001) and the difficulty of designing compatible multiplex primer sets (Elnifro et al., 2000). Moreover, differentiation of many serotypes is a labour-intensive procedure and subject to ambiguous results.

The progress assessed in microarray technology and its combination with PCR permits to perform a more specific, sensitive and robust detection of viruses (Chen et al., 2006; Gheit et al., 2006; Hwang et al., 2003; Lapa et al., 2002). Replacing gel electrophoresis analysis of the PCR products with DNA–DNA hybridization reduces the ambiguous detection of nonspecific sequences that might be present within amplified products. Moreover, hybridization assays allow to use an increased number of probes for detection of multiple human pathogenic viruses (Chizhikov et al., 2002; Hwang et al., 2003; Kawaguchi et al., 2003; Lapa et al., 2002; Lin et al., 2004; Wang et al., 2002). Unfortunately, probe design is always a high-time consuming task due to the huge amount of genomic sequences available in public databases. This requires dealing with hundreds or thousands of DNA sequences, which renders manual screening impracticable. In this study, a novel and rapid approach to simplify and accelerate the processing of all necessary sequences

required to perform this type of experiments is described. This new technology has been successfully applied for the detection and serotyping of adenoviruses causing acute respiratory disease. This approach permits to design probes for the microarray assays as well as validate PCR primers for the selected sequences avoiding cross-reactions among them.

In summary, an oligonucleotide microarray assay has been developed to detect and identify simultaneously adenovirus serotypes causing respiratory tract infection by using a new tool for the retrieval of sequences between two previously defined flanking regions or primers. This microarray assay is capable of identifying and serotyping simultaneously the most common adenovirus serotypes 1, 2, 3, 4, 5, 7, 14, and 21 associated with adult respiratory tract infections and childhood acute respiratory disease in laboratory and clinical specimens with efficiency and accuracy. This method has the advantages of being easier and less expensive than others previously described. Besides, it could easily be extended to any other pathogen diagnosis and can be implemented in laboratories with few facilities and limited resources.

2. Materials and methods

2.1. Prototype adenovirus strains and controls

Reference strains of ADV serotypes 1, 2, 3, 4, 5, 7, 14, and 21 were used as positive controls. Reference strains of other 49 ADV serotypes were used as control of specificity. To assess the sensitivity of the microarray detection, additional controls were prepared in three dilutions that contained 10, 100 and 1000 molecules of cloned amplified product of the ADV serotype 1, as well as each type of influenza virus (A, B, C) and respiratory syncytial virus (RSV A and B). These controls were used to determine the specificity of the method and they were described previously (Coiras et al., 2003). The number of copies per tube was adjusted after measurement of the OD₂₆₀ from plasmid DNA purified with Wizard Plus SV Miniprep DNA Purification System (Promega, Madison, WI). The primary amplification product from the ADV serotype 1 cloned in pGEM-T vector was previously described (Avellón et al., 2001). Negative controls with RNase-free sterile water (Sigma, St. Louis, MO) were included in each batch of test samples for checking up carryover contamination.

2.2. Clinical specimens

Nasopharyngeal aspirates from children presenting acute respiratory syndromes as bronchiolitis and pneumonia, or an influenza-like illness, were submitted from the Severo Ochoa Hospital, in Madrid, Spain. These clinical specimens were received for virological study at the Respiratory Virus laboratory, in the National Centre of Microbiology (CNM, Instituto de Salud Carlos III, Majadahonda, Madrid, Spain). All specimens were collected in 3 ml virus transport medium (MEM, Gibco BRL, Life Technologies, Paisley, Scotland; penicillin 200 U/ml and streptomycin 200 µg/ml, BioWhittaker, Walkersville, MD; mycostatin 200 U/ml, Sigma–Aldrich, St. Louis,

MO; bovine serum albumin 0.25%, Merck, Darmstadt, Germany). The homogenized sample was divided in aliquots to perform IFI and PCR assays to identify the most important respiratory viruses, including adenovirus. Their isolation in the appropriate cell lines was also performed. Semi-confluent monolayers of Madin–Darby canine kidney (MDCK), human laryngeal epidermoid carcinoma (HEp-2) cells, and human lung diploid (HEL) cells were inoculated with 200 μ l of homogenized samples to detect the presence of influenza viruses, RSV and ADV. MDCK cells were maintained in Eagle's minimal essential medium with penicillin–streptomycin supplemented with 3 μ g/ml of trypsin. HEp-2 and HEL cells were maintained in Eagle's minimal essential medium supplemented with 2% fetal bovine serum, 200 UI/ml penicillin, and 0.1 mg/ml streptomycin.

2.3. PCR assay and labeling

Nucleic acids were extracted from 200 μ l of the sample using the guanidinium thiocyanate extraction method previously described (Casas et al., 1995). Negative controls were treated following the same procedure. After processing, the dried pellet was resuspended in 15 μ l of RNase-free sterile water (Sigma–Aldrich). Nucleic acids extracted from the controls and the clinical samples were amplified using generic primers for adenovirus that were designed inside the hexon protein-coding region, described previously (Avellón et al., 2001), by using a nested PCR assay described in Coiras et al. (2003). Briefly, PCR amplification reaction was performed in a PCR mixture containing 2 mM $MgSO_4$, 200 μ M each of dATP, dGTP, dCTP, and dTTP, 20 pmol of specific primers, 10 μ l of 5 \times reaction buffer, and 5 U of DNA polymerase (Promega, Madison, WI). A 5 μ l aliquot of DNA extracts was added to a final volume of 50 μ l. Amplifications were carried out into thin-walled reaction tubes (Sorenson, BioScience, UT) in a PTC-200 (Peltier Thermal Cycler, MJ Research, Watertown, MA). Samples were subjected to the following cycling conditions: 45 cycles, 94 °C for 30 s; 50 °C for 1 min; 68 °C for 1 min, and a final incubation of 68 °C for 10 min. A 2 μ l aliquot from this first reaction was added to a final volume of 50 μ l to perform fluorescent Cy3 labeling of the samples by nested PCR assay. Nested PCR mixture contained 2 mM $MgCl_2$ (Perkin Elmer, Branchburg, NJ), 200 μ M each of dATP, dGTP, dTTP (Amersham Pharmacia Biotech, Piscataway, NJ) and Cy3-dCTP:dCTP 1:9 (GE Healthcare, Uppsala, Sweden), 20 pmol of each specific primer, 60 mM Tris–HCl (pH 8.5), 15 mM $(NH_4)_2SO_4$, and 1.25 U AmpliTaq DNA Polymerase (Perkin Elmer). Before PCR, samples were heated to 95 °C for 4 min. Cycling conditions were as follows: 35 cycles, 94 °C for 30 s; 55 °C for 1 min; 72 °C for 30 s, and a final incubation of 72 °C for 10 min.

2.4. Oligonucleotide probe design

The internal primers of the nested PCR were used to target and specifically amplify the hexon protein coding-region. Therefore, the probes for microarray assay were selected within this region to perform a generic detection of human adenoviruses

Table 1

Oligonucleotide probes designed for the detection of the main adenovirus serotypes involved in acute respiratory disease

Probes	Sequence	Length	T_m (°C)
gADV	TTYAACCAYCACCGYAAYGCKGG	23	77.3
ADV-a	TTGCTGGGCAACGGTCGCTACGTGC	25	83.9
ADV-b	TGTTGTTGGGAAACGGCCGC	19	75.1
ADV-c	GCTGGTCTGCGCTACAGGTCCATGC	25	83.9
ADV-d	GCTACGTGCCATTCCACATCCAGGT	25	80.6
ADV-e	TCTCCTGCCGGGCTCATACACCTAC	25	82.3
ADV-f	CTGGYCTACGCTRCGGTCCATGCT	25	82.3
ADV-g	CCGCCATCTCTAGTAGACACCTATG	25	79
ADV-h	GCTTTTGGGCAATGGTCGTTACGTG	25	79

Each adenovirus serotype could be recognized by the generic probe (gADV) and by more than one probe (see text).

as well as specific identification of the most important adenoviruses involved in respiratory infection: serotypes 1, 2, 3, 4, 5, 7, 14, and 21 (see Table 1). The selection and retrieval of those internal amplicons among all the sequences of human adenovirus available in Genbank was performed by using BussuB software, described below. Those selected regions were then aligned using ClustalX software (Thompson et al., 1997) to detect the conserved regions and select only those where the variability among the different genotypes was observed. The oligonucleotides were analyzed using Oligo6 software (Molecular Biology Insights, Inc., Cascade, CO) and checked afterwards with the Basic Local Alignment Search Tool (BLAST) of the National Center for Biotechnology Information (NCBI) to avoid cross-reactivity. All the probes were designed to have similar physical–chemical properties to ensure that they were able to hybridize under the same conditions without modifying their target affinity.

2.5. BussuB amplicon retrieval software

Microarray probe design is a complex process that requires a deep knowledge of the primary sequences to be used. In the present approach, these probes were designed to detect amplicons generated by a previously published nested PCR assay (Coiras et al., 2003). All the human adenovirus' sequences available at the moment were recovered from Genbank and analyzed prior to probe design. Using BussuB tool, all possible virtual amplicons generated with these primers were retrieved and aligned to detect variable regions susceptible of being used as probes.

BussuB is an amplicon retrieval software implemented using JAVA technology (<http://www.java.sun.com>) in two different versions: a web-based one (<http://ibsen.tic.udc.es/bussuWeb2>) and a command-line stand-alone version, available upon request. BussuB can be used to detect all the possible amplicons generated from two given primers on a provided set of FASTA format sequences. BussuB input parameters include: (1) FASTA format file with the sequences from where subsequences would be retrieved; (2) both primers or flanking regions that limit the subsequences of interest to be retrieved; (3) number of bases to be considered as part of the 3' end control region; (4) num-

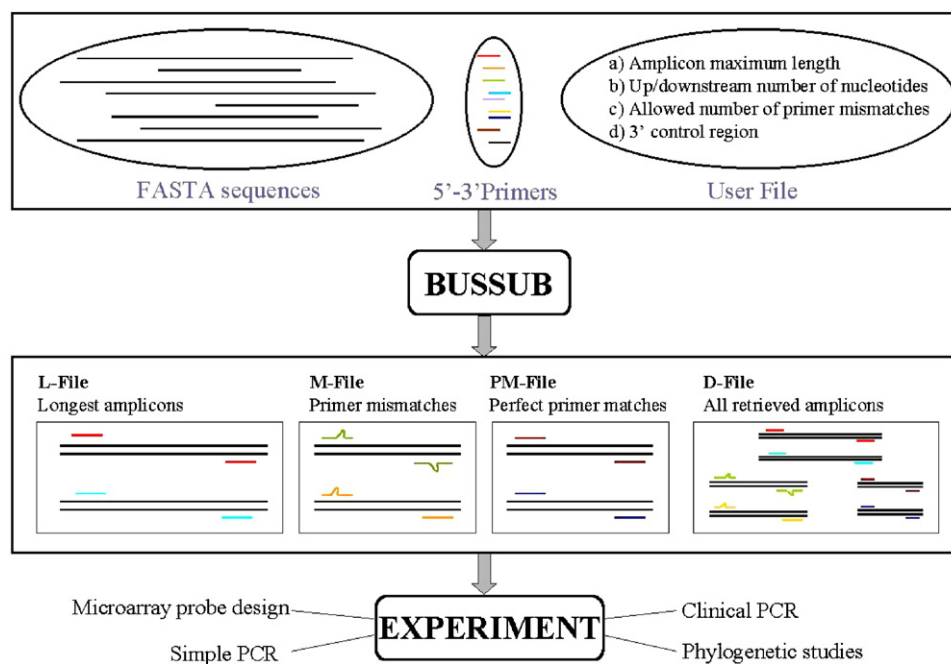


Fig. 1. Schematic representation of the BussuB performance. The upper box contains the necessary input data to use this tool: all sequences to be analyzed in FASTA format; the 5' and 3' primers to amplify the PCR product; and four restriction parameters. After the BussuB algorithm processes this information, four output files are generated (included in the lower box). These files (L, M, PM and D) contain relevant information that will be used to develop further experiments, such as microarrays, PCR assays or phylogenetic studies (see text).

ber of allowed mismatches within the primers; (5) maximum size of the expected amplicons; and (6) length in bases of the sequences to be retrieved upstream or downstream from the identified primers or defined flanking regions. Once all the input parameters have been defined, the algorithm scans sequentially the input FASTA file extracting from each sequence all possible amplicons according to the parameters defined by the user (Fig. 1). The output consists on four different files: L-file, containing the longest retrieved subsequences that perfectly match conditions for the primers/flanking regions; M-file, a FASTA format file containing the retrieved subsequences according to the limits previously defined by the user, including one or more mismatches; PM-file, a FASTA format file containing the subsequences according to the limits previously defined by the user, with perfect matching in the primers/flanking regions; and D-file, a tabular formatted file including all the retrieved sequences (with and without mismatches) including a brief description such as retrieved sequence length and GC%. In the case of the sequences retrieved with mismatches, the number of mismatches is also included as a descriptor in this file.

2.6. Slide functionalization and microarray printing

Oligonucleotides were printed on poly-l-lysine derived microscope glass slides. Functionalization of slides was performed as described previously (Scheda et al., 1995). Briefly, microscope slides (Erie Scientific, Portsmouth, NH) were washed for 2 h in a cleaning solution (NaOH:H₂O:EtOH 1:4:6 m/v/v) and then rinsed several times in ddH₂O before

transferring them to the poly-l-lysine solution (poly-l-lysine: PBS1×:H₂O 1:1:8 v/v/v) (Sigma-Aldrich) for the functionalization step. Slides were kept under gentle shaking for 60 min and then washed four times with ddH₂O. The slides were dried by centrifugation and baked for 1 h at 45 °C. Once functionalized, slides were kept at room temperature for 1 week before printing. Microarray printing was performed by using an Affymetrix-417 Arrayer (Affymetrix, Santa Clara, CA). In brief, the oligonucleotide probes were resuspended at 40 μM and 10 μM in 96-well microtiter plates using 3× SSC as printing buffer. Each probe was printed in replicates in different regions of the functionalized slide. Once the slides were spotted they were baked for 1 h at 85 °C to improve the probe immobilization to the surface.

2.7. DNA–DNA hybridization

Five microliters of the Cy3-labeled samples obtained in the nested PCR assay were mixed with hybridization buffer (final concentration 5× SSC, 10% formamide, 0.1%SDS) and denatured for 99 °C for 5 min followed by and incubation of 5 min at 4 °C. Then, SDS was added after denaturing rising to a final volume of 14 μl and sample was dispensed onto the microarray surface. The slide was covered with a glass cover-slip, enclosed in a Genotek microarray hybridization chamber (Genotek, Barcelona, Spain) and then hybridized for 2 h at 42 °C in a hybridization oven. After hybridization, microarrays were washed 5 min in SSC 1×/0.1%SDS, then 5 min in SSC 0.1×/0.1%SDS, and rinsed with SSC 0.1× and 95%EtOH before spinning dry.

2.8. Microarray scanning and image analysis

Dried microarrays were scanned using an Affymetrix-418 scanner (Affymetrix). The generated images were quantified and analyzed using the ImaGene software (Biodiscovery, Inc., Sunnydale, CA). Data were stored and analyzed in an in-house MIAME-compliant database (López-Campos et al., 2006).

2.9. Sequencing and phylogenetic analysis

The amplified fragments obtained by nested PCR were subjected to electrophoresis in 2% Seakem agarose gels (FMC) in Tris–borate buffer containing 0.5 µg/ml ethidium bromide and subsequently purified with the QIAquick PCR Purification Kit (Qiagen, Hilden, Germany). Nucleotide sequences were determined using the BigDye Terminator Cycle Sequencing v. 2.0 Ready Reaction Kit (Applied Biosystems, Foster City, CA). The primers used for sequencing were those used in the nested PCR at 3.2 pmol per 100 ng DNA template, per reaction. Phylogenetic analysis of the sequences was performed in order to identify the genotype of the adenoviruses present in the clinical specimens. To this end, the sequences of the hexon protein coding-region amplified by the nested PCR assay were analyzed by Molecular Evolutionary Genetics Analysis 3.1 software (MEGA) (Kumar et al., 2004) with 10,000 bootstrap replicates.

3. Results

3.1. Labeling of PCR products

A nested PCR assay described previously (Coiras et al., 2003) was employed to define the region of the adenovirus genome used as target for the microarray assay. The selected primers that amplified partially the hexon protein coding-region allowed the detection of any adenovirus serotype (Avellón et al., 2001). Within this region, a probe for generic detection of any human adenovirus was designed, as well as specific probes for detecting and identifying the most prevalent adenoviruses involved in human respiratory infection: 1, 2, 3, 4, 5, 7, 14, and 21 serotypes.

The nested PCR assay was performed to obtain Cy3-labeled samples and to provide enough material for the further hybridization process. The fluorescent labeling of the amplicon was performed during the second amplification step. A non-significant decrease in the efficiency of the PCR reaction due to the incorporation of the dye was observed (data not shown). In case of the PCR products used for sequencing and phylogenetic analysis, no Cy3-labeling was performed in order to avoid interference with the sequencing dye.

Hybridization results were independent of whether the PCR assay products used had been previously purified or not (data not shown). Therefore, all subsequent hybridization assays were performed using directly the PCR labeled products without any further processing.

3.2. Probe design by using BussuB software

Probes were designed to identify specifically the most important adenoviruses involved in respiratory infection: serotypes 1, 2, 3, 4, 5, 7, 14, and 21 (see Table 1). Oligonucleotide probe design was improved by using the BussuB software. This tool permits recovering the sequence contained within two flanking regions by processing large quantities of DNA sequences with very high accuracy. Besides, it has been designed to be able to recognize and accept IUPAC code for degenerations. Another important aspect inherent to viruses is viral variation and mutation. In order to address this process, BussuB allows the user to define a number of mismatches that can occur in the flanking regions. These changes will be unknown a priori and are different to those defined by the degenerated bases in the primers. Moreover, BussuB offers the possibility of defining a 3' end polymerase specificity region. This parameter is determined by the user in the input file as the number of bases within the 3' end of the primers which cannot contain any mismatch. Therefore, the user can define specifically how many bases must be recognized by the polymerase in the 3' end of the primer. Any other position in the primer will be suitable to be changed and considered as a possible mutation that may affect PCR efficiency but not avoid primer elongation. Finally, the upstream and downstream parameters available in BussuB input have the aim of retrieving external regions to those defined by the primers/flanking regions. Due to these characteristics, BussuB might be applied in different environments, such as phylogenetic analysis or virtual PCR screening (Fig. 2).

To ensure BussuB reliability, several validation trials were performed. First, multi-sequence input files were used to test the system computational capacity. Once this step was successful, a real biological problem was tested. A clinical PCR for detecting dengue viruses was selected (Holmes and Burch, 2000; Rico-Hesse, 2003). This model was chosen due to its complexity – since RNA viruses present a high variability in their sequences – to check BussuB's use of degenerate primers and mismatch options. As a result, BussuB was able to detect all the expected sequences due to its capability to work with degenerate primers. It also detected some other amplification products that could be seen in gel electrophoresis but were not expected from BLAST analysis due to viral sequence variation (data not shown).

Subsequently, a clinical PCR assay for detecting human adenoviruses was selected (Coiras et al., 2003) because these viruses are divided in 52 serotypes that can only be differentiated by slight differences within their hexon protein coding-region sequences. Since adenoviruses are DNA viruses, the primers showed a high level of specificity.

3.3. Oligonucleotide microarray design

The oligonucleotide microarray for detection of adenoviruses was designed not only to detect any adenovirus present in the sample by using a generic probe (gADV), but also to identify simultaneously the most important adenovirus serotypes involved in human respiratory infection—1, 2, 3, 4, 5, 7, 14, and 21 serotypes – by using specific probes (Table 1). To perform

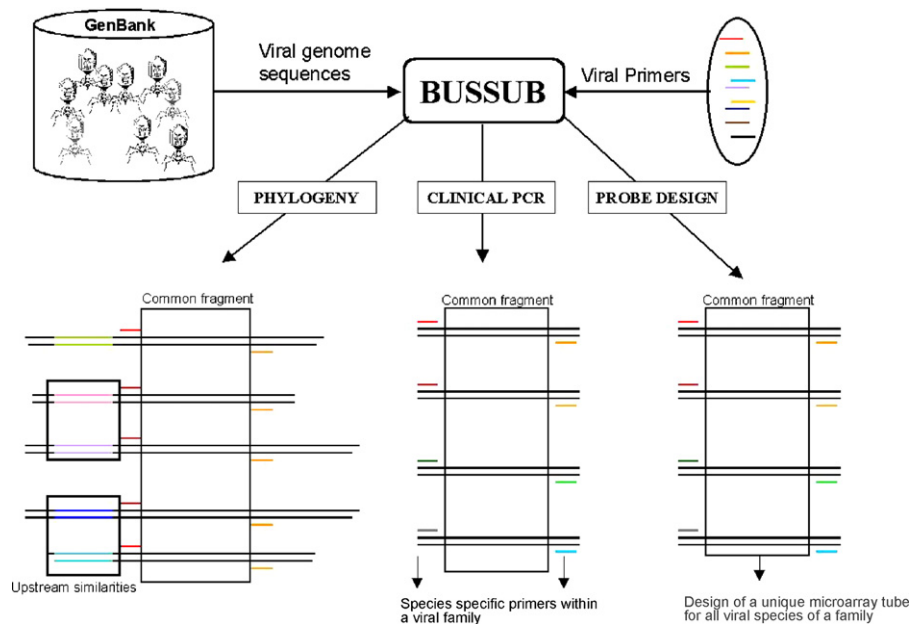


Fig. 2. BussuB applications: BussuB produces four output files that may be used in different types of experiments—(a) phylogeny: the upstream region of a sequence could detect phylogenetic relationships between genomic sequences of a viral family; (b) clinical PCR: a pair of primers can discriminate between different sequences of the same viral family; (c) probe design for hybridization assays: if different sequences of the same viral family match within a region defined by a pair of primers, a unique probe could be designed within this region to be able to detect all the viruses included in this family.

a more accurate detection, these probes were designed to obtain a characteristic hybridization pattern for every adenovirus serotype (Fig. 3). As a result, a particular and distinguishable hybridization pattern with the immobilized probes was obtained for every virus that permitted the simple visual recognition of the adenovirus serotype present in the clinical specimen.

3.4. Specificity and sensitivity of the assay

The specificity of these probes was tested with the whole genome of reference adenovirus strains 1–49 and no significant cross-reaction with other adenoviruses different from the serotypes 1, 2, 3, 4, 5, 7, 14, and 21 was found. Nucleic acids from other respiratory viruses such as influenza virus types A, B and C, and respiratory syncytial virus types A and B were used as control of specificity and no signal was detected in any case (data not shown). Sensitivity of the microarray assay was performed by using controls that contained 10, 100 and 1000 molecules of cloned amplified product of the ADV serotype 1 as well as different dilutions of the reference strains of adenoviruses from the serotypes 1, 2, 3, 4, 5, 7, 14, and 21. Microarray-based sensitivity assessments revealed lower detection limits for adenovirus generic detection (between 10 and 100 genomic copies) and serotyping (between 100 and 1000 genomic copies) in both clinical specimens and prototype strains.

3.5. Validation of the microarray assay with clinical specimens

The validity of the microarray assay was tested for the genotyping of clinical samples. Ten clinical specimens coming from nasopharyngeal aspirates already classified as adenoviral infec-

tions by PCR, IFI or isolation in cell culture, as described above, were used to be genotyped using the microarray assay. After scanning the slides, hybridization images were assigned to the different genotypes on the basis of the image layout. A second analysis was carried out by a simple quantification of the signal intensity coming from the spots in comparison to the previously established numerical patterns for every respiratory genotype (Fig. 4A). All adenoviruses detected in these specimens were sequenced to perform a phylogenetic analysis and compare the results with those obtained in the microarray (Fig. 4B). The phylogenetic tree was generated using a neighbor joining inference method with 10,000 bootstrap replicates. In all cases, the genotyping of the serotypes tested with the microarray assay matched with the serotype assignment by sequencing and phylogeny. Separate clinical specimens infected with other respiratory viruses – such as RSV or influenza – were used as control of specificity. No hybridization was detected when the virus present in the sample was not an adenovirus (data not shown).

4. Discussion

The rapid development of new technologies based on molecular biology has allowed for the improvement of infectious disease diagnosis. A wide variety of new methods has gradually replaced the conventional assays used in routine diagnosis, and has led to a high interest in developing assays for detection and typing of human respiratory viruses. Methods such as PCR/RT-PCR (Avellón et al., 2001; Faix et al., 2004; Na et al., 2002; Papapetropoulou and Vantarakis, 1998; Pring-Akerblom et al., 1999; Xu and Erdman, 2001; Xu et al., 2000), sequencing (Takeuchi et al., 1999), or microarrays (Lin et al., 2004), or their combinations (Chen et al., 2006; Gheit et al., 2006; Hwang et

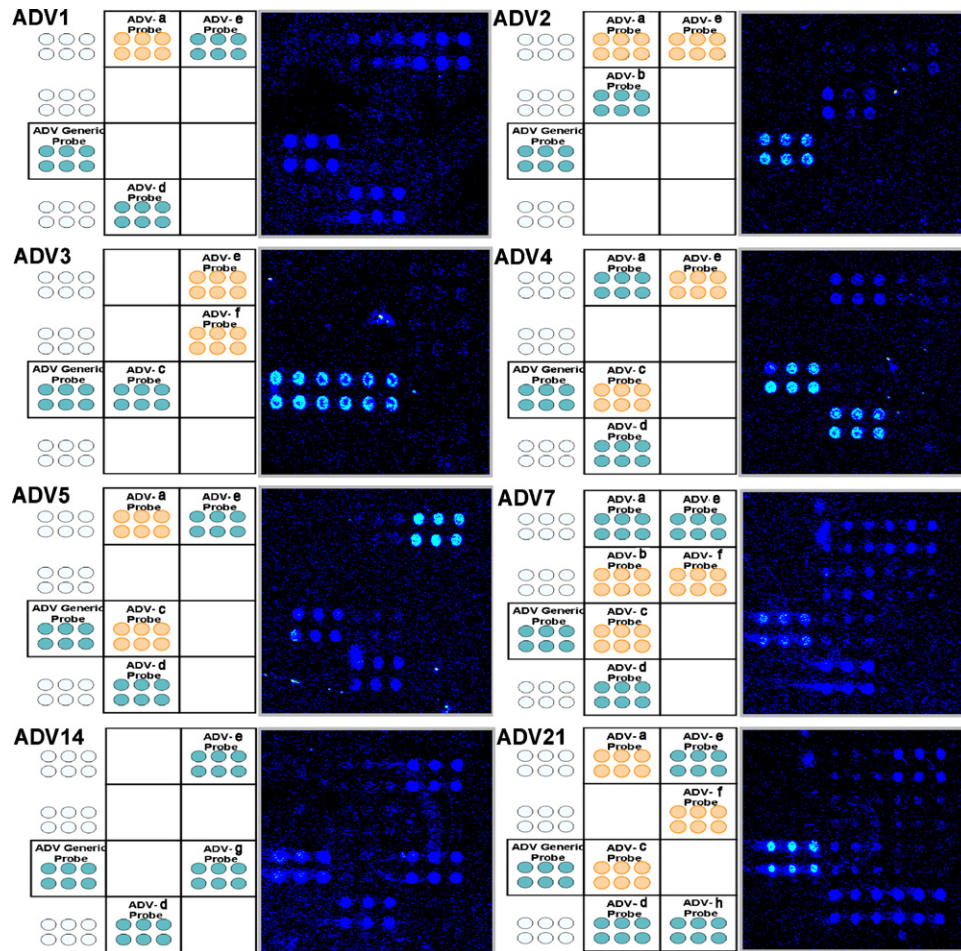


Fig. 3. Characteristic hybridization patterns of the adenovirus reference strains used to standardize the microarray assay. The images are quite clean even considering that the hybridization was performed using directly the PCR products without further purification. Adenovirus type 1 was detected by probes ADV-a, -d and -e; adenovirus type 2 was detected by probes ADV-a, -b and -e; adenovirus type 3 was detected by probes ADV-c, -e and -f; adenovirus type 4 was detected by probes ADV-a, -c, -d and -e; adenovirus type 5 was detected by probes ADV-a, -c, -d and -e (with different intensities); adenovirus type 7 was detected by probes ADV-a, -b, -c, -d, -e and -f; adenovirus type 14 was detected by probes ADV-d, -e and -g; and adenovirus type 21 was detected by probes ADV-a, -c, -d, -e, -f and -h.

al., 2003; Lapa et al., 2002) have the advantage of improving speed and sensitivity over culture-based procedures and other classic assays but tend to be expensive and technically demanding. However, these methods could be applied progressively in clinical diagnosis for rapid identification of virus causing infection in order to determine the right therapy, prophylaxis or isolation procedures. Therefore, the simplification of such molecular methods is necessary to improve the routine diagnosis in hospitals.

For this purpose, a combined nested PCR and microarray-based approach for the simultaneous detection and genotyping of adenoviruses causing human acute respiratory disease has been developed and standardized. Adenovirus infection is endemic, particularly among children, but also may cause epidemic of pharyngoconjunctival fever, keratoconjunctivitis, gastroenteritis and acute respiratory disease in crowded communities. The absence of effective treatment for acute respiratory disease makes essential the development of a diagnostic method for rapid detection and identification of the causing viruses and taking the adequate control measurements to stop the transmission of infection. However, since the adenoviridae family includes

numerous serotype groups, probe design could be a highly time-consuming task. This requires dealing with hundreds or thousands of DNA sequences for which manual screening is unfeasible. To cope with them, an automatic amplicon retrieval software that simplifies and accelerates the processing of all sequences necessary to perform this type of experiments has been developed. This software, BussuB, allows for the analysis of huge amount of genomic sequences available in public databases in order to design oligonucleotides for PCR assays, microarrays or phylogenetic studies (Figs. 1 and 2). In fact, BussuB offers the possibility of obtaining all possible amplified sequences from primers with a determined number of mismatches. The primers obtained can be easily defined by the user and are highly specific for the selected viral gene. BussuB is able to deal with high complexity files including complete bacterial genomes, providing fast and accurate results. The efficiency of the system is clearly superior compared to manual methods when a large number of sequences are required for primer design. BussuB has some advantages when compared to other tools previously described, such as: permitting the use of IUPAC code for degenerate bases within the primers; upstream and downstream

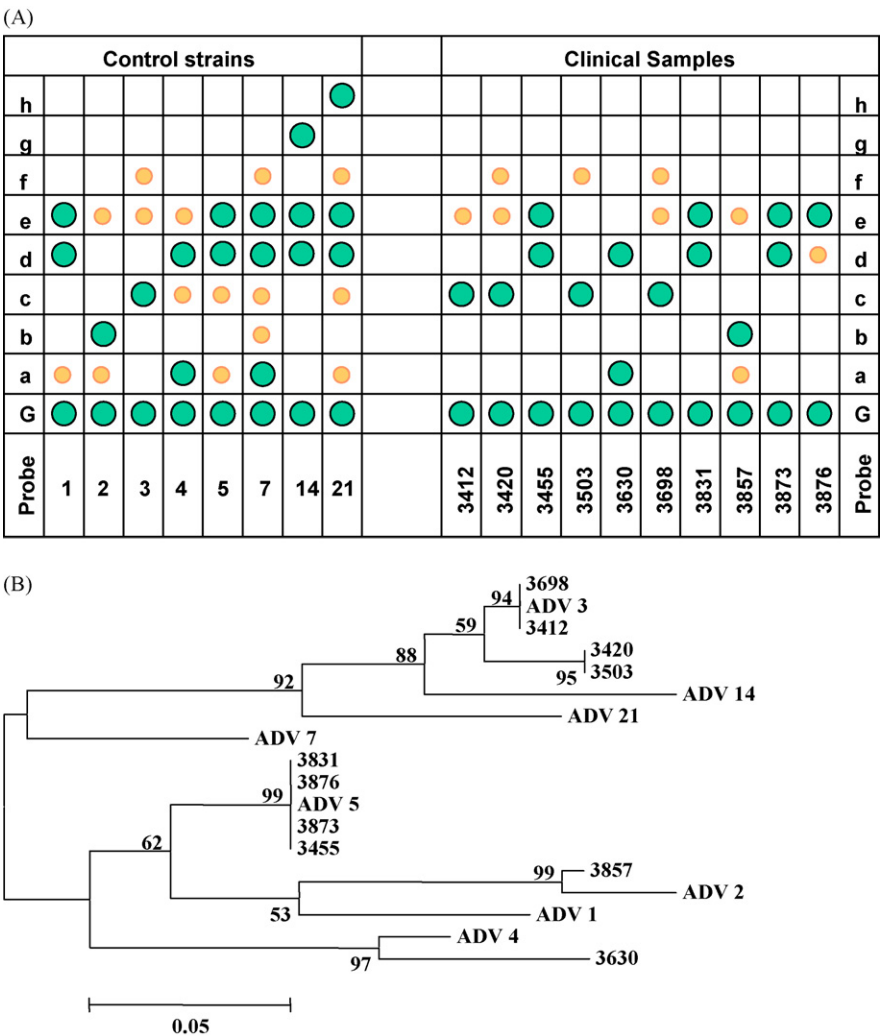


Fig. 4. Validation of the microarray assay with clinical specimens: (A) hybridization patterns obtained from the measured intensities of the spots considered as positives for each sample. The different hybridization intensities are represented with different diameters and colors being the big green circles the ones with higher hybridization intensities and the small orange those ones with lower hybridization intensities and (B) phylogenetic analysis of the clinical samples used for the validation of the method. Sequences of different respiratory adenoviruses obtained from the GenBank were used as reference to assign the genotype for each adenovirus present in the clinical specimens.

selection of primer/flanking regions; possibility of application to cross-reactivity analysis or probe design and also for phylogenetic studies by selecting contiguous regions to conserved sequences. These analyses could facilitate the tracking of a viral infection origin or any other evolutionary aspect. FASTA formatted output files also facilitates the use of the results from these analysis with other bioinformatics tools.

In the present work, a microarray assay for adenovirus detection and serotyping is described. This method is able to simultaneously detect and identify the most important adenovirus serotypes involved in acute respiratory disease – 1, 2, 3, 4, 5, 7, 14, and 21 – by using generic and specific probes (Table 1). These probes were located within the amplified hexon-protein coding-region and they were Cy3-labeled by a nested PCR assay. The microarray design allows a visual identification of the serotype detected by showing a characteristic hybridization pattern for each adenovirus (Fig. 3). The sensitivity of the assay for adenovirus generic detection was determined as 10–100

molecules and 100–1000 molecules for specific serotyping. The specificity was also tested using the reference adenovirus strains 1–49, excluding the serotypes 1, 2, 3, 4, 5, 7, 14, and 21, as well as other respiratory viruses like influenza viruses or RSV. No hybridization was detected in any case (data not shown).

To validate the relevance of this microarray assay for simultaneous detection and identification of the adenovirus serotypes causing acute respiratory disease in a clinical environment, ten clinical specimens consisting in nasopharyngeal aspirates already classified as adenoviral infections by other methods were used. To confirm the results obtained, sequencing and phylogenetic analysis were performed with the same samples (Fig. 4). In all cases, the results of genotyping obtained by the microarray assay matched with those obtained by sequencing. As expected, all adenoviruses detected had tropism for the respiratory tract. Besides, the detection of serotypes 2, 3, 4 and 5 showed the most representative adenoviruses that circulate currently within Spanish population (Casas et al., 2005).

Another goal of this work was to obtain a diagnosis system that could be easily used in facilities with limited resources. This has been clearly achieved by the method proposed in this paper, since only a scanner for the detection of the hybridization results is required. Chemicals and number of processing steps necessary for the experiments have also been minimized. Thus, the number of steps has been reduced to just four: (1) sample extraction; (2) sample amplification and labeling; (3) hybridization without a previous purification step; and (4) scanning and simple data interpretation.

In conclusion, an easy to implement oligonucleotide microarray has been developed to simultaneously detect and identify adenovirus serotypes causing respiratory tract infection with efficiency and accuracy. The design of a characteristic hybridization pattern that permits a simple identification of the adenovirus serotype as well as the simplification of the hybridization conditions makes this method a good candidate for rapid clinical diagnosis. In addition, new retrieval software BussuB has been developed to simplify the identification of the best target DNA sequences. This software was able to determine the gene region within DNA or RNA genome where the probes must be designed to be used in a microarray system with high specificity and sensitivity.

Patent interests statement

The authors declare the existence of patent protection for the oligonucleotide probe gADV described in Table 1 (Patent Application P200403083).

Acknowledgements

The authors thank the Spanish Surveillance Network for providing the clinical specimens. Reference adenovirus serotypes were kindly provided by Dr. Juan Echevarría (Centro Nacional de Microbiología, Instituto de Salud Carlos III, Madrid, Spain). The primary amplification product from the adenovirus serotype 1 cloned in pGEM-T vector was kindly provided by Dr. Ana Avellón (CNM, ISCH).

References

- Allard, A., Albinsson, B., Wadell, G., 2001. Rapid typing of human adenoviruses by a general PCR combined with restriction endonuclease analysis. *J. Clin. Microbiol.* 39, 498–505.
- Avellón, A., Pérez, P., Aguilar, J.C., Ortiz de Lejarazu, R., Echevarría, J.E., 2001. Rapid and sensitive diagnosis of human adenovirus infections by a generic polymerase chain reaction. *J. Virol. Met.* 92, 113–120.
- Broude, N.E., Driscoll, K., Cantor, C.R., 2001. High-level multiplex DNA amplification. *Antisense Nucleic Acid Drug Dev.* 11, 327–332.
- Casas, I., Powell, L., Klapper, P.E., Cleator, G.M., 1995. New method for the extraction of viral RNA and DNA from cerebrospinal fluid for use in the polymerase chain reaction. *J. Virol. Met.* 53, 25–36.
- Casas, I., Avellón, A., Mosquera, M., Jabado, O., Echevarría, J.E., Campos, R.H., Rewers, M., Perez-Brena, P., Lipkin, W.I., Palacios, G., 2005. Molecular identification of adenoviruses in clinical samples by analyzing a partial hexon genomic region. *J. Clin. Microbiol.* 43, 6176–6182.
- Chen, T.C., Chen, G.W., Hsiung, C.A., Yang, J.Y., Shih, S.R., Lai, Y.K., Juang, J.L., 2006. Combining multiplex reverse transcription-PCR and a diagnostic microarray to detect and differentiate enterovirus 71 and coxsackievirus A16. *J. Clin. Microbiol.* 44, 2212–2219.
- Chizhikov, V., Wagner, M., Ivshina, A., Hocino, Y., Kapikian, A.Z., Chumakov, K., 2002. Detection and genotyping of human group A rotaviruses by oligonucleotide microarray hybridization. *J. Clin. Microbiol.* 40, 2398–2407.
- Coiras, M.T., Perez-Brena, P., Garcia, M.L., Casas, I., 2003. Simultaneous detection of influenza A, B, and C viruses, respiratory syncytial virus, and adenoviruses in clinical samples by multiplex reverse transcription nested-PCR assay. *J. Med. Virol.* 69, 132–144.
- De Jong, J.C., Wermenbol, A.G., Verweij-Uijterwaal, M.W., Slaterus, K.W., Wertheim-Van Dillen, P., Van Doornum, G.J., Khoo, S.H., Hierholzer, J.C., 1999. Adenoviruses from human immunodeficiency virus-infected individuals, including two strains that represent new candidate serotypes Ad50 and Ad51 of species B1 and D, respectively. *J. Clin. Microbiol.* 37, 3940–3945.
- Elnifro, E.M., Ashshi, A.M., Cooper, R.J., Klapper, P.E., 2000. Multiplex PCR, optimization and application in diagnostic virology. *Clin. Microbiol. Rev.* 13, 559–570.
- Faix, D.J., Houg, H.S., Gaydos, J.C., Liu, S.K., Connors, J.T., Brown, X., Asher, L.V., Vaughn, D.W., Binn, L.N., 2004. Evaluation of a rapid quantitative diagnostic test for adenovirus type 4. *Clin. Infect. Dis.* 38, 391–397.
- Gheit, T., Landi, S., Gemignani, F., Snijders, P.J., Vaccarella, S., Franceschi, S., Canzian, F., Tommasino, M., 2006. Development of a sensitive and specific assay combining multiplex PCR and DNA microarray primer extension to detect high-risk mucosal human papillomavirus types. *J. Clin. Microbiol.* 44, 2025–2031.
- Grondahl, B., Puppe, W., Hoppe, A., Kuhne, I., Weigl, J.A., Schmitt, H.J., 1999. Rapid identification of nine microorganisms causing acute respiratory tract infections by single-tube multiplex reverse transcription-PCR, feasibility study. *J. Clin. Microbiol.* 37, 1–7.
- Holmes, E.C., Burch, S.S., 2000. The causes and consequences of genetic variation in dengue virus. *Trends Microbiol.* 8, 74–77.
- Horwitz, M.S., 2001. Adenoviruses. In: Knipe, D.M., Howley, P.M. (Eds.), *Fields Virology*, 4th ed. Lippincott/The Williams & Wilkins Co., Philadelphia, PA, pp. 2301–2326.
- Hwang, T.S., Jeong, J.K., Park, M., Han, H.S., Choi, H.K., Park, T.S., 2003. Detection and typing of HPV genotypes in various cervical lesions by HPV oligonucleotide microarray. *Gynecol. Oncol.* 90, 51–56.
- Kawaguchi, K., Kaneko, S., Honda, M., Kawai, H.F., Shiota, Y., Kobayashi, K., 2003. Detection of hepatitis B virus DNA in sera from patients with chronic hepatitis B virus infection by DNA microarray method. *J. Clin. Microbiol.* 41, 1701–1704.
- Kumar, S., Tamura, K., Nei, M., 2004. MEGA3 integrated software for molecular evolutionary genetics analysis and sequence alignment. *Brief. Bioinform.* 5, 150–163.
- Lapa, S., Mikheev, M., Shchelkunov, S., Mikhailovich, V., Sobolev, A., Blinov, V., Babkin, I., Guskov, A., Sokunova, E., Zasedatelev, A., Sandakchiev, L., Mirzabekov, A., 2002. Species-level identification of orthopoxviruses with an oligonucleotide microchip. *J. Clin. Microbiol.* 40, 753–757.
- Lin, B., Vora, G.J., Thach, D., Walter, E., Metzgar, D., Tibbetts, C., Stenger, D.A., 2004. Use of oligonucleotide microarrays for rapid detection and serotyping of acute respiratory disease-associated adenoviruses. *J. Clin. Microbiol.* 42, 3232–3239.
- López-Campos, G., Garcia-Albert, L., Martin-Sanchez, F., Garcia-Saiz, A., 2006. Analysis and management of HIV peptide microarray experiments. *Met. Inf. Med.* 45, 158–162.
- Mitchell, L.S., Taylor, B., Reimels, W., Barrett, F.F., Devincenzo, J.P., 2000. Adenovirus 7a, a community-acquired outbreak in a children's hospital. *Pediatr. Infect. Dis. J.* 19, 996–1000.
- Na, B.K., Kim, J.H., Shin, G.C., Lee, J.Y., Lee, J.S., Kang, C., Kim, W.J., 2002. Detection and typing of respiratory adenoviruses in a single-tube multiplex polymerase chain reaction. *J. Med. Virol.* 66, 512–517.
- Papapetropoulou, M., Vantarakis, A.C., 1998. Detection of adenovirus outbreak at a municipal swimming pool by nested PCR amplification. *J. Infect.* 36, 101–103.
- Poddar, S.K., 1999. Detection of adenovirus using PCR and molecular beacon. *J. Virol. Met.* 82, 19–26.

- Pring-Akerblom, P., Trijssenaar, F.E., Adrian, T., Hoyer, H., 1999. Multiplex polymerase chain reaction for subgenus-specific detection of human adenoviruses in clinical samples. *J. Med. Virol.* 58, 87–92.
- Rico-Hesse, R., 2003. Microevolution and virulence of dengue viruses. *Adv. Virus Res.* 59, 315–341.
- Saitoh-Inagawa, W., Oshima, A., Aoki, K., Itoh, N., Isobe, K., Uchio, E., Ohno, S., Nakajima, H., Hata, K., Ishiko, H., 1996. Rapid diagnosis of adenoviral conjunctivitis by PCR and restriction fragment length polymorphism analysis. *J. Clin. Microbiol.* 34, 2113–2116.
- Schena, M., Shalon, D., Davis, R.W., Brown, P.O., 1995. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science* 270, 467–470.
- Schmitz, H., Wigand, R., Heinrich, W., 1983. Worldwide epidemiology of human adenovirus infections. *Am. J. Epidemiol.* 117, 455–466.
- Shenk, T.E., 2001. Adenoviruses. In: Knipe, D.M., Howley, P.M. (Eds.), *Fields Virology*, 4th ed. Lippincott/The Williams & Wilkins Co., Philadelphia, PA, pp. 2265–2300.
- Takeuchi, S., Itoh, N., Uchio, E., Aoki, K., Ohno, S., 1999. Serotyping of adenoviruses on conjunctival scrapings by PCR and sequence analysis. *J. Clin. Microbiol.* 37, 1839–1845.
- Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F., Higgins, D.G., 1997. The ClustalX windows interface, flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res.* 24, 4876–4882.
- Van der Veen, J., Dijkman, J.H., 1962. Association of type 21 adenovirus with acute respiratory illness in military recruits. *Am. J. Hyg.* 76, 149–159.
- Wang, D., Coscoy, L., Zylberberg, M., Avila, P.C., Boushey, H.A., Ganem, D., DeRisi, J.L., 2002. Microarray-based detection and genotyping of viral pathogens. *Proc. Natl. Acad. Sci. U.S.A.* 99, 15687–15692.
- Willcox, N., Mautner, V., 1976. Antigenic determinants of adenovirus capsids. II. Homogeneity of hexons, and accessibility of their determinants, in the virion. *J. Immunol.* 116, 25–29.
- Xu, W., Erdman, D.D., 2001. Type-specific identification of human adenovirus 3, 7 and 21 by a multiplex PCR assay. *J. Med. Virol.* 64, 537–542.
- Xu, W., McDonough, M.C., Erdman, D.D., 2000. Species-specific identification of human adenoviruses by a multiplex PCR assay. *J. Clin. Microbiol.* 38, 4114–4120.

Addressing the Biomedical Informatics Needs of a Microarray Laboratory in a Clinical Microbiology Context

Guillermo LOPEZ-CAMPOS, Victoria LOPEZ ALONSO¹, Fernando MARTIN-SANCHEZ

*Medical Bioinformatics Department National Institute of Health “Carlos III”,
Madrid, Spain*

Abstract. For an effective integration of microarray-based technologies in clinical settings a number of contributions from biomedical informatics technologies and techniques are needed to facilitate the improvement of the phases of experimental design, image analysis, data management, annotation, and analysis. In this communication we briefly present the state-of-the-art in the application of biomedical informatics to laboratories conducting microarray experiments and how our unit is coping with these requirements imposed by the routine clinical work of the National Centre of Microbiology, a reference laboratory for the Spanish Health System.

Keywords: Biomedical informatics, microarray, bioinformatics clinical microbiology.

Introduction

Microarray technology is being integrated into basic biomedical research and is becoming a fundamental molecular monitoring tool in clinical microbiology settings, especially for diagnostic applications due to the possibility of multiplexing the simultaneous detection of several pathogens in a single reaction [1, 2, 3]. Most of these clinical applications are based in the development of DNA microarrays for the detection of specific genes or gene regions in pathogens, but others are based in protein (ELISA microarrays) or other immobilised molecules. Clinical microbiologists can also apply microarray technology from other approaches such as analysis of gene expression to monitor microbial metabolism and screening of regulons to study microbial response to drugs, environmental changes, genome organization, and evolutionary studies [4].

Many software development projects have emerged in response to the needs of postgenomic research projects focused on laboratory information management [5], scripting and programming language (bioperl, biopython, biojava, bioruby...) [6], database integration [7] and data models and ontology development [8].

Along with this increasing diversity of laboratory techniques and uses of microarrays there is an increasing need of specific biomedical informatics tools in some areas of the field of microarray applications in clinical microbiology that are not covered by the

¹Victoria López Alonso, Institute of Health “Carlos III”, victorialopez@isciii.es

solutions designed and developed for gene expression studies, mainly in cancer research.

The aim of this work is to identify some of the biomedical informatics needs detected in a microarray laboratory focused on clinical microbiology (Figure 1) and how these user requirements could benefit from different synergistic developments of bioinformatics and medical informatics techniques and methods [9].

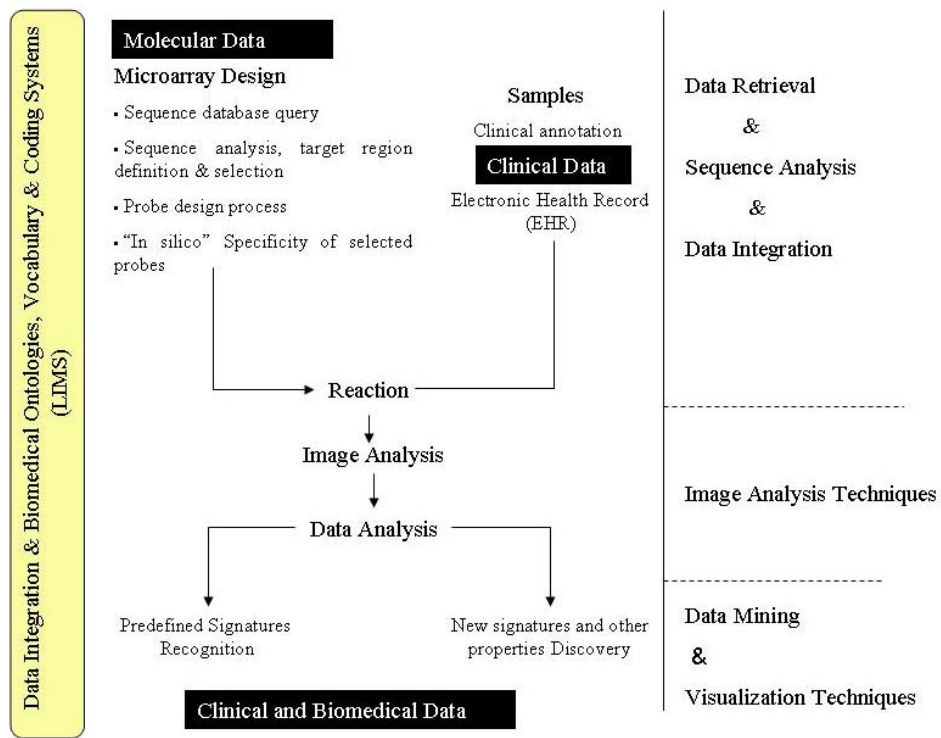


Figure 1. Biomedical informatics technologies of application in the different phases of a microarray experiment in a clinical microbiology context.

1. Material and Methods

1.1. Biomedical informatics and Microarray Design

Technological advances have boosted the field of microarray experiments. Although they represent high-throughput methods microarrays must be carefully designed in order to obtain useful data for clinical applications. Therefore design steps are key processes for the success or failure in the routine use of microarrays. In most of the clinical microbiological applications of microarrays, the design process is focused on the detection of probes that are capable to characterize and specifically identify the organism or organisms of interest. Therefore the initial step of design consists of

querying sequence databases and retrieving information. The target is the sequence of DNA to be amplified and fluorescently labelled. The strategy used to design it involves interrogating complete genome sequences stored in specific databases as the Comprehensive Microbial Genome [10] and expressed sequence tag databases available in the public domain. This step is problematic due in some cases to the vast amount and high redundancy of available sequence data or in other cases to the lack of available sequences in databases. At this point considerable effort has already been spent in generating EST libraries for the design of microbial genome microarrays.

Microarray probe design tools are very inefficient for large genomes and many existing tools operate in a batch mode. There are now several technologies that are useful for the design of microarrays for a selected set of model organisms based on Gene Ontology (GO) classification or using UniGene clusters [11].

Specificity of the designed oligonucleotides has to be checked with BLAST (Basic Local Alignment Search Tool) [12] and smart filtering techniques, which are employed to avoid redundant computation while maintaining accuracy.

1.2 Data management and annotation for Clinical Microarrays

Managing a microarray laboratory, especially for clinical use, typically involves good organization and sound tracking of experiments and samples data, including information on clinical samples, treatment, and experimental results. The use of Laboratory Information Management Systems (LIMS) addresses this problem by storing information with rigorous data entry mechanisms to prevent inaccuracies and to avoid redundancy. LIMS represent repositories of raw data associated with experiments, and are extremely useful as transient repositories before final submission of experimental microarray results to public databases, such as Gene Expression Omnibus (GEO) [13] or ArrayExpress [14].

1.3. Image Analysis Techniques in Microarray Experiments

In medical image processing the image content is often represented by features that are computed from the pixel matrix in order to extract features that support the development of reliable clinical diagnostic systems. In the case of microarray images there are some added difficulties for image analysis because many features are of abstract nature, as for instance those derived from a wavelength transform function.

After a reaction, microarrays are scanned and image analysis software determines the raw values using a data file that identifies the features and defines their dimensions and locations. Data stored in a tagged image file format (TIFF) is quantified by the image analysis software. To quantify microarray images it is necessary to ensure the quality of the signals and to prevent artefacts. Finally, the arithmetic mean or better median of the foreground and background pixels is calculated and data is obtained in the form of a table [15].

1.4. Microarray Data Mining Techniques

Before obtaining microarray results and applying sophisticated statistics, it is important to adequately normalize the data. A number of normalization methods have been proposed to correct systematic biases resulting basically from different amounts of

DNA used for labelling, different efficiencies of the Cy-3 and Cy-5 dyes in the labelling protocols, and different detection efficiencies of the dyes [16]. In gene expression experiments, fundamental patterns are extracted by several clustering methods like hierarchical clustering, self organizing maps and support vector machines. Lastly, the use of additional knowledge sources in the microarray data analysis process can improve the discovery and validation of clinical hypothesis. Information on sequence and structure, gene and protein interactions, function annotation and ontologies, or genetics and metabolic pathways can significantly complement any data analysis and improve its results.

2. Results

2.1. BUSSUB, a tool for designing clinical microarrays

The use of oligonucleotide microarray technology requires a very detailed attention to the design of specific probes to be spotted on the solid phase. They must have high sensitivity to detect a target gene and must be unique with high specificity. Parameters such as probe length, number of oligonucleotides, maximal distance from the 3 end, melting temperature range, threshold to reject secondary structures and prohibited sequences have to be considered in order to obtain an efficient design. In addition, the designed oligonucleotides must be computationally optimized to achieve greater specificity and uniformity and to perform optimally under the same melting temperature and other experimental conditions, reducing the noise due to cross-hybridization. Although there is an increasing number of publicly available genetic and molecular analysis programs for microarray oligonucleotide design, most of them use the same algorithm or criteria with only scarce variation [17]. None of them gathers the whole set of criteria. For this reason we have developed BUSSUB [3] an amplicon retrieval software that simplify and boost the process of recovering sequences contained between two given regions. BUSSUB is essential is the first stage of microarray design because this system is able to deal with high complexity files including several complete bacterial genomes.

2.2. AMANDA, Clinical Microarray Laboratory Information Management System

We have developed AMANDA [2] a LIMS specifically designed to store information about patient data, sample description, experimental resources, experimental parameters and conditions, and raw and processed hybridization results. The integration of this information into the microarray information system facilitates the processes of quality control and assessment of microarray experiments as well as further clinical interpretation of the results. AMANDA is compliant with the minimal information about a microarray experiment (MIAME) format [18]. This system is able to store, manage and visualize all the data and the information coming from several studies, to process it as a homogeneous dataset instead of multiple and separate sources

2.3 Image Analysis Software

Image analysis software used in our microarray laboratory carries out three fundamental tasks: gridding (to locate each spot on the slide), segmentation (to

differentiate the pixels within a spot-containing region into foreground) and information extraction. Information extraction includes two steps: spot intensity extraction and background intensity extraction. Spot intensity extraction refers to the calculation of fluorescent signal of the foreground from segmentation process, while background intensity extraction utilizes different algorithms to estimate the background signal due to the non-specific hybridization on the glass.

Developments in image acquisition, analysis and informatics technologies are ongoing and are expected to broaden the usefulness of DNA microarrays. For example Stanford MicroArray Database (SMD) [19] has the ability to store, retrieve, display and analyze the complete raw data produced by several additional microarray platforms and image analysis software.

2.4. Information of Clinical Microarray experiments

Most of the information resulting from diagnostic microarray experiments in a clinical microbiology context is qualitative (presence or absence of signal). In other cases as in gene expression or protein assays, data need to be analyzed in the context of clinical and epidemiological information in order to extract relevant knowledge useful for developing clinical solutions (Figure 2).

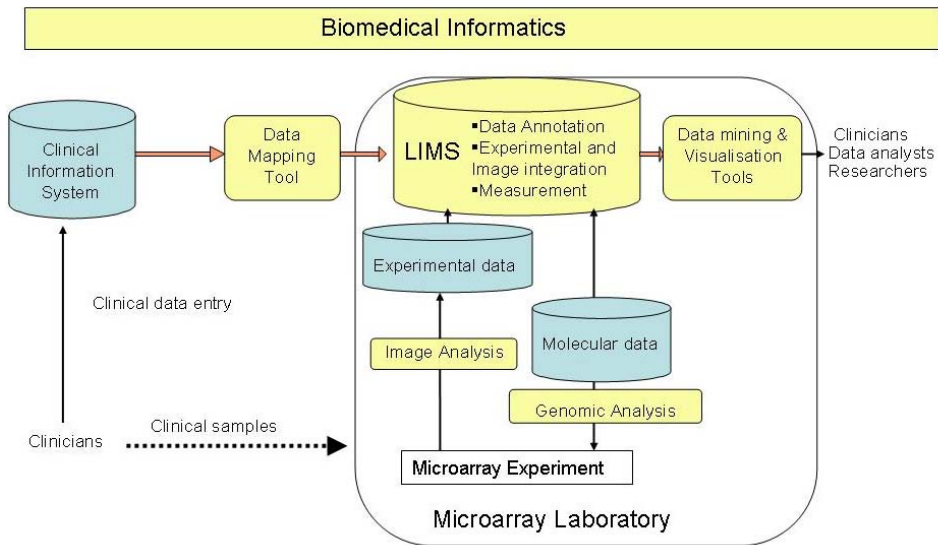


Figure 2. Implementation of biomedical informatics tools required for the production and use of clinical microarrays.

Data arising from microarray experiments can be normalized, filtered and analyzed utilizing several normalization and statistical modules available on integrative web based services. SMD [19] and the Bioconductor project [20] have added several tools and new ontologies to allow for an accurate and searchable annotation of biological

samples and experiments, developing and generalizing the schema for more efficient and flexible storage and analysis of microarray data.

Conclusions

High-throughput technologies such as microarrays are in the process of revolutionizing clinical microbiology. Further research and development of new biomedical informatics methods and techniques is becoming an integral part of a microarray laboratory. In our experience, these technologies are key in the process of assigning experimental genomic signatures to clinical profiles and to turn data into meaningful and reproducible clinical and mechanistic inferences.

References

- [1] A. Loury and L. Bodrossy, "Highly parallel microbial diagnostics using oligonucleotide microarrays", *Clinica Chimica Acta*, vol. 363, pp. 106-119, 2006.
- [2] G. López-Campos, L. García-Albert, F. Martín Sanchez and A., García-Saiz, "Analysis and management of HIV peptide microarray experiments", *Methods of Information in Medicine*, vol 45, pp. 158-162, 2006.
- [3] G. López-Campos, M. Coiras, J.P. Sánchez-Merino, M.R. López-Huertas, I. Spiteri, F. Martín-Sánchez, P. Pérez-Breña, "Oligonucleotide microarray design for detection and serotyping of human respiratory adenoviruses by using a virtual amplicon retrieval software" *Journal of Virological Methods*, vol. 145, pp. 127-136, 2007.
- [4] A. Ehrenreich, "DNA microarray technology for the microbiologist: an overview" *Applied Microbiology and Biotechnology* vol. 73, pp. 255-273, 2006.
- [5] G. Thallinger, S. Trajaneski, G. Stocker and Z. Trajanoski, "Information management systems for pharmacogenomics", *Pharmacogenomics*, vol 3 pp. 651-667, 2002.
- [6] <http://www.open-bio.org>
- [7] V. Maojo, M. García-Remesal, H. Billhardt, R. Alonso-Calvo, D. Perez-Rey and F. Martín-Sánchez, "Designing new methodologies for integrating biomedical information in clinical trials", *Methods of Information in Medicine*, vol 45, pp 180-185, 2006.
- [8] <http://www.obofoundry.org/ontologies.shtml>
- [9] D. Rebholz-Schuhman, G. Cameron, D. Clark, E. van Mulligen, J.L. Coatrieux, E. Del Hoyo Barbolla, F. Martín-Sánchez, L. Milanesi, I. Porro, F. Beltrame, I. Tollis and J. Van del Lei, "SYMBIOmatics: Synergies in Medical Informatics and Bioinformatics-exploring current scientific literature for emerging topics", *BMC Bioinformatics*, vol 8, (Suppl 1) S18, 2007.
- [10] <http://cmr.tigr.org/tigr-scripts/CMR/CmrHomePage.cgi>
- [11] "Onto-Tools, the toolkit of the modern biologist: Onto-Express, Onto-Compare, Onto-Design and Onto-Translate", *Nucleic Acids Research*. vol 31, pp 3775-3781, 2003.
- [12] <http://www.ncbi.nlm.nih.gov/BLAST/>
- [13] T. Barret and R. Edgar, "Mining microarray data at NCBI's Gene Expression Omnibus (GEO)", *Methods of Molecular Biology*, vol. 338, pp. 175-190, 2006.
- [14] H. Parkinson, M. Kapushesky, M. Shojatalab, N. Abeygunawardena, R. Coulson, A. Farne, E. Holloway, N. Kolesnykov, P. Lilja, M. Lukk, R. Mani, T. Rayner, A. Sharma, E. William, U. Sarkans and A. Brazma, "ArrayExpress-a public database of microarray experiments and gene expression profiles", *Nucleic Acids Research*, vol. 35, pp. 747-750.
- [15] A. Petrov, S. Shah, S. Draghici, S. Shams, "Microarray image processing and quality control" in S. Shah, G. Kamberova (eds) *DNA array image analysis- nuts & bolts*. DNA, Eagleville, pp 99-130
- [16] J. Quackenbush, "Microarray data normalization", *Nature Genetics*, vol 32, pp 496-501
- [17] W. Rychlik "OLIGO7" *Methods of Molecular Biology*, vol. 402, pp 35-60, 2007.
- [18] A. Brazma, P. Hingamp, J. Quackenbush, G. Sherlock, P. Spellman, C. Stoecker, J. Aach, W. Ansorge, C.A. Ball, H.C. Causton et al. "Minimum information about a microarray experiment (MIAME) - toward standards for microarray data", *Nature Genetics*, vol 29, pp 365-371, 2001.
- [19] <http://genome-www5.stanford.edu/resources/restech.shtml>
- [20] <http://www.bioconductor.org>